Mental Workload in Aviation: An Investigation of Physiological and Qualitative Assessment Methods

Author:
Ridwan Aji Budi Prasetyo

Supervisors:
Dr Robert Houghton
Dr Elizabeth Argyle

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Institute for Aerospace Technology
&
Human Factors Research Group
Faculty of Engineering
The University of Nottingham

December 2022

Abstract

Flying is a safety critical activity in which the ability of the pilot to synthesise multiple sources of information, make decisions and produce appropriate control inputs is critical. Consequently, understanding and managing pilots' mental workload (MWL) is critical to flying safely and in the design of flight-decks and procedures. Investigation of objective ways to measure MWL is necessary as recent advances in sensing technology offer new opportunities to develop assessment methods that are less intrusive than existing techniques. For example, physiological methods are emerging as options for MWL assessment as they can provide *in-situ* measurement. These methods have potential advantages in terms of being relatively less intrusive than traditional methods, offering benefits for real workplace settings, including in the cockpit. The research presented in this thesis focuses on the exploration of MWL measurement during a simulated flying task through the investigation of factors influencing MWL during flight and the utility of using more objective methods for evaluating it.

Four studies were conducted. First, acceptance of real-time mental workload monitoring was explored among pilots and passengers using a combination of interviews, surveys and online methods. A Critical Decision Method (CDM) interview was then applied to professional pilots to understand the factors that influence pilots' experiences of high mental workload during a flight. Two connected experiments were undertaken to test the utility of physiological sensors for detecting changes in mental workload during a simulated flying task. Finally, an online experiment was undertaken to evaluate vicarious estimation of workload by human observers based on videos of task performance.

This combination of studies contributes to the understanding of physiological measurements of MWL during a simulated flying task. Firstly, this thesis provides insights about professionals' and public attitude towards MWL sensors technology in the future. Secondly, this thesis offers further understanding of what makes pilots experience high mental workload during landing, and it can be used as the basis for the development of a simulated flying task. Thirdly, this thesis contributes to supporting the spatial resolution of the fNIRS. More specifically, the thesis suggests that the left-side of the prefrontal cortex was activated in response to MWL changes during the simulated flying task. The notion that heart rate measures and pupil dilation can indicate MWL changes were also supported by this study. Finally, this thesis offers an initial understanding that MWL during a simulated flying task cannot be accurately predicted before the task, unless the contrasting elements of the task can be shown.

Acknowledgements

All praises and thanks are only for Allah, the One who, by His blessing and favour, perfected goodness.

To begin with, I wish to thank my supervisors, Dr Robert Houghton and Dr Elizabeth Argyle, for their endless guidance at each stage of my research; also to Dr Paula Moran and Prof Sarah Sharples for their roles as my supervisor during the first two-year of my PhD programme. Thanks also go to Marie Sklodowska-Curie Actions and The University of Nottingham for sponsoring this PhD programme.

I also want to thank everyone in Marie-Curie INNOVATIVE programme, Institute for Aerospace Technology, and Human Factors Research Group at the University of Nottingham. In particular, I would like to thank Dr Adrian Marinescu and Dr Horia Maior for their technical guidance during my research. Thanks also go to everyone who has directly helped my research, i.e. the pilots and the members of the public.

Furthermore, I would also like to thank my Indonesian friends and communities in Nottingham (PPI, IndoSoc, and PeDLN) for their warm friendship and, of course, delicious foods. In particular, special thanks go to the member of 'Rumah Kapitalis' (Mr. R. F. Aditra and Mr. A. W. Mahdiyasa) for providing me temporary settlement during my thesis writing period. Special thanks also go to Psychology Department at Brawijaya University for liaising with my employment status during this PhD programme.

Finally, I wish to thank all of my family. In particular, I want to express my gratitude to my beautiful wife, Nourmaya Masyitha, for her love and constant support, for all the late nights and early mornings, and for keeping me sane over the past few years. However, keep in mind that The Bodyshop is overpriced, go to Aldi instead for skincare! For the two biggest blessings of my life: R. A. Prasetyo and M. M. Prasetyo, I hope I can always be a good father for you. Moreover, the completion of this PhD would not be possible without support from my parents; I thank you for providing me a good life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Research Background and Motivation

Mental workload (MWL) is a topic of enduring interest in human factors and ergonomics. The concept arguably emerged during the 1970s, formally communicated in 1976 during a conference sponsored by NATO (Moray, 1979). However, the emergence of the concept can also be traced back to research topics during World War II which focused on the operation and safety of warplanes, exploring themes such as fatigue, human error, task demands, design of the cockpit, and pressurisation (Waterson, 2011). The concept of MWL has arguably become ever more prominent in recent times since automation and digitalisation have changed many tasks and roles such that cognitive work is emphasised over physical work.

Aviation is a field where MWL research can have serious safety implications for operations. While previous research has explored MWL experienced during flight (e.g. Wilson, 2002; Dahlstrom et al., 2011) and the domain has offered a setting for developing theories of workload (e.g. Helleberg & Wickens, 2003; Edwards et al., 2012), what is less well-understood is what generates MWL during a real flight and how it can be evaluated using more objective methods. This thesis therefore explores the concept of MWL in aviation, focusing on filling the gap in knowledge surrounding effective MWL evaluation with emerging methods. Furthermore, the research explores perceptions of MWL assessment technology, pilot decision-making during a critical flying phase, physiological response to MWL changes during a simulated flying task, and MWL prospective prediction. These endeavours provide an improved understanding of MWL in the sense of linking the actual MWL experienced by pilots with more objective MWL assessment methods, and in addition, the social aspect of the potential future technology's implementation complements the explanation. This is important to achieve the objective of advancing understanding of MWL in flight and the development of viable future assessment technologies.

MWL is also a concept with strong intuitive sense, tallying with everyday lived experience, and it appears to be common in many work environments. For example, in everyday work tasks we may feel that we have to attend to too many items of information, or perhaps when driving, we find that focusing on the road means we cannot pay attention to a discussion with a passenger. However, reaching agreement on its definition, both technically and philosophically, is surprisingly difficult (Young et al., 2015; Linton et al., 1989). Several definitions have been offered, emphasising the interplay between task demands (Sharples & Megaw, 2015), human cognitive or attention processing capacity (Kantowitz, 2000), operator performance (Hart & Staveland, 1988), subjective experiences (Van Acker et al., 2018), and most recently, the way the brain works (Ayaz et al., 2012). However, the definition of MWL seems pragmatic and relative to the context. It follows the operational aspect of the tasks under question, particularly when "it comes to application of tools in real-world settings" (Sharples, 2019, pp. 491). In summary, a universal, commonly accepted definition of MWL is non-existent. The definition of MWL appears to be dependent on the area of application or perhaps on each individual researcher (Cain, 2007). For the purposes of this research, MWL is defined and understood as the interaction between task demands, performance or behavioural response, and subjective experiences from internal and external sides of a person (Sharples & Megaw, 2015). This definition was considered appropriate for this research since it incorporates those three main aspects of MWL frequently discussed in the literature.

Piloting an aircraft involves many inherently complex tasks, with the potential for generating high levels of MWL. The nature of flying an aircraft requires the pilot to perform multiple tasks concurrently. Common terms among aviators, namely 'to aviate', 'to navigate', and 'to communicate', indicate the principal group of tasks that a pilot should perform to achieve a safe flight. The ordering of phrases also indicates the hierarchy of task importance (Schutte & Trujillo, 1996), which means that tasks related to controlling the aircraft ('to aviate') should be prioritised over 'to navigate' or 'to communicate' tasks in an overload or emergency. The multitasking environment that pilots encounter therefore has consequences on mental processes when performing their jobs.

Understanding the demands imposed by a task and its effects on operators' limited cognitive resources is considered essential to achieve an improved working environment, a more intuitive work station design, or more effective procedures (Cain, 2007).

Several methods for measuring MWL have been proposed, with the trend in recent research shifting to the utilisation of physiological sensors to capture physiological changes when performing a task. Among the physiological sensors, the more inexpensive and practical options may be promising in operational contexts, with various techniques such as functional near infrared spectroscopy (fNIRS) (Foy et al., 2016), heart rate measures (Mansikka et al., 2016), eye-gaze behaviour (Marquart et al., 2015), facial thermography (Marinescu et al., 2018), and brain activity through blood oxygenation (Ayaz et al., 2012).

Physiological techniques have a great advantage in terms of being less intrusive than probe-based or observational methods, giving a potential for use in real workplace settings. The stream of data obtained from physiological devices could also potentially be processed in real time to provide feedback to operators (Maior et al., 2018). These two potential benefits, low intrusiveness and real-time feedback, are essential regarding the practicality of workload measurement. They also offer opportunities to advance mental workload research, particularly regarding the potential implementation of the methods in real work settings.

In an aircraft itself, technology is available to support pilot decision-making in the form of sensors that provide warnings about physical aspects of a flight. For example, the Traffic Collision and Avoidance Warning System (TCAS/TCAWS) aims to alert pilots about surrounding traffic. If another aircraft is flying on a collision course with an initial aircraft, the system will provide auditory and visual alerts, along with suggestions on actions pilots must follow (e.g. 'descend!'). The Ground Proximity Warning System (GPWS) works similarly, providing pilots with feedback and suggested actions if the aircraft is flying into the ground or an obstacle.

However, all of these sensors are related to the aircraft itself or to the environment, but not to the pilot's own performance. Physiological measures are one potential resource that could be translated into wearable sensors that can capture changes in pilots' mental workload. There potentially would be considerable benefits from identifying changes in pilots' mental workload, or simply said: how cognitively 'busy' the pilots are while flying the aircraft. This may benefit the aviation world even further, particularly in safety through workload management, especially during crucial phases of a flight. The feedback it provides might also support the implementation of widely used Crew Resource Management (CRM) concept in the cockpit and thus improve safety. This is considered important as humans are prone to errors. Human factors such as situation awareness and non-adherence to procedure contribute to most incidents in aviation and these factors tend to emerge unconsciously, indicating the pressure in the cockpit (Kharoufah et al., 2018). Moreover, the implementation of Single-Pilot Operation (SPO) in the future appears to be likely to save operational costs without jeopardising safety. The conceptual framework of SPO has, in fact, been proposed, emphasising the needs for automation capability improvements in the cockpit (Bilimoria et al., 2014). One of the automation capabilities that requires development to support SPO is pilot health monitoring, and it can be achieved by applying physiological sensors to the pilot-in-command. With physiological sensors, more objective MWL quantification and MWL self-monitoring could be made possible for supporting a safe SPO.

Investigations to achieve the goal of MWL monitoring in flight therefore need to be advanced, particularly by conducting research in more contextual environments. In the context of aviation, studying MWL may advance the use of physiological sensors in a simulated flying task or a flight simulator by real professional pilots. It may also support the concept of the 'contextual digital footprint' Sharples & Houghton (2017) for performance evaluation since the concept emphasises the utilisation of data from many digital systems over a certain period of time. For instance, data from physiological sensors can be linked to the data from aircraft behaviour to generate more comprehensive insights about pilot performance. This might be applied in the aviation context gradually, from a simulator to the real flight, and using various types of sensors, from the more complicated types, such as EEG or fNIRS, to the more practical types, such as smartwatches.

The focus of this research, therefore, is to improve the understanding of mental workload during a flying task and predict it using novel measurement methods. This focus is achieved by investigating professional and public acceptance toward the concept of MWL sensor technology implementation and by understanding the dynamics of pilots' mental workload in real-world settings. Moreover, achieving the focus of the research is also conducted by investigating the way physiological methods for MWL measurement could capture MWL changes in a simulated and continuous flying task, and lastly, by developing initial understanding of prospective prediction of MWL in a flying task. This thesis contributes to the understanding of MWL during a flying task through the investigation of methods for assessing MWL using qualitative and physiological approaches. This research assumes that there is a connection between mental state, physiological response, and behavioural response. This connection will serve as the logical basis for the research presented in this thesis. Mental workload can be considered as the mental state of a person when responding to certain stimuli (Wulvik et al., 2020). This mental state is believed to be reflected in specific physiological response from certain organs in the body, such as the brain, the heart, and the eyes. It is also assumed that behavioural response will emerge when responding to the stimuli with the state of mental and physiological response.

1.2 Thesis Research Question, Aims, and Overview

The thesis investigates the research question *"Can we improve our understanding of mental workload during a flying task and evaluate it using novel measurement methods?"* To address this question, we developed four research aims:

1. To understand MWL assessment methods in the literature and opportunities for using physiological measures to estimate MWL in flying tasks. We begin the project by exploring previous studies regarding the concept of mental workload. We focused our review on essential aspects of MWL, namely the definition, underlying theories, MWL measurement framework and methods, and current issues regarding MWL. We also reviewed a potential task environment that can resemble a flying task with different levels of MWL. The result from reviewing the literature was the decision to use several physiological sensors to detect changes in MWL during a flying task, simulated by a task battery.

2. Examine attitudes of professionals and public regarding a hypothetical scenario for the implementation of workload sensor technology in future cockpits. To begin the empirical study for this research project, we started by gathering attitudes from pilots and members of the public with a hypothetical scenario involving the implementation of MWL sensors in the cockpit. A study consisting of three activities was undertaken, incorporating both qualitative and quantitative methods. The first activity was an online-survey of professional pilots regarding their attitudes toward the hypothetical scenario of MWL sensors implementation in the cockpit. The second activity deepened the response from the first activity through an online Focus Group Discussion (FGD) interview. Meanwhile, the third activity targeted members of the public for their attitudes towards the implementation using an online experiment. From the results of this study, most surveyed pilots agreed with the implementation. The FGD interview revealed several issues that were likely to shape their positive attitudes, such as potential comfort, versus negative attitudes, such as invalidity of the sensors.

3. Explore cognitive processes of pilots in an operational flying task and identify sources of task demands during a flying task. Upon learning that the pilots and members of the public supported the implementation of MWL sensors, testing the potential devices was our next step. However, we needed to learn what actually makes a pilot cognitively ‘busy’ during a flying task so that we could identify the elements of it and understand factors influencing MWL generation. To this end, we employed a Critical Decision Method (CDM) interview technique to obtain richer data regarding pilots' cognitive processes during a flying task, particularly during a landing phase. The landing phase was chosen as it has been considered to be the most crucial phase during a flight regarding safety. This phase, therefore, was more likely to contain various and dynamic cognitive processes and thus task demands. The results from this study provided potential sources of task demands and their interaction with performance and external/internal influence, creating dynamic MWL during landing.

4. Develop tasks that resemble flying task demands and investigate MWL measurements using physiological methods. This research aim was developed to explore potential physiological sensors such as brain sensors (fNIRS), heart rate sensors, and pupillary devices to detect changes in MWL during a flying task. To increase the validity of the measurement, we developed a task from the Multi-Attribute Task Battery (MATB) environment to resemble task demands during a flying task. The task consisted of tracking, managing fuel balance, and monitoring the system; it was chosen by reviewing the literature and considering the results from the previous aim (no. 3). Two lab experiments were conducted with various levels of task demands to represent a real flying experience. With different levels of task demand, mental workload is expected to differ; thus, various levels of brain activation and physiological response will be detected. The results from this study revealed that fNIRS sensors were insufficiently sensitive for detecting MWL changes during the task, whilst partial heart indicators and pupillary measures seemed to be sufficiently sensitive.

5. Examine if MWL can be predicted before a task. This aim was extended from the previous research aim, focusing on if we can detect changes in MWL during a flying task merely by assessing the description and a sample of the task. An online experiment was applied to examine participants' responses to subjective MWL of a flying task by carefully observing the videos of the task. The results from this study revealed that MWL during MATB task cannot be predicted prospectively yet partial elements constituting the task might be. This study provided an initial understanding of prospective MWL prediction during a flying task.

The studies addressed the aforementioned questions and aims are presented in this thesis in the following manner, as shown in Table 1.1. The table provides an overview of studies carried out, including their purpose, methods used, and the relevant chapters in which they are discussed.

Table 1.1: An overview of studies in the thesis.

Study	Aim	Methods Employed	Chapter
n/a	To understand MWL assessment methods in the literature and opportunities for using physiological measures to estimate MWL in flying tasks.	Literature review	2
1	To examine attitudes of pilots and public towards the implementation of MWL sensors technology in future cockpit.	Online survey, focus group discussion, online experiment	3
2	To explore cognitive processes of a real flying experiment from subject-matter experts (SMEs) and identify sources of task demands during a flying task.	Critical decision method (CDM) interview	4
3	To develop task that can resemble task demands of a flying task and investigate the ability of physiological measures in detecting MWL changes during a flying task.	Literature review, lab experiment	2, 5
4	To investigate whether MWL changes during a flying task can be predicted before the task.	Online experiment	6

1.3 Scope of the Research

The research in this thesis is designed to understand MWL during a flying task and novel assessment methods, particularly physiological techniques. Therefore, the intent of this study is to investigate MWL changes exclusively during a task in the aviation context. The physiological techniques used in this research explored three measures, namely, brain haemodynamic response using fNIRS, heart rate (HR) using a HR monitoring device, and pupil dilation using an eye-tracker. The selection of the sensors was based on the results from previous studies, with some pragmatic considerations that were also considered. Concerning participants for the studies, whilst we used real professional pilots for the survey and FGD in Study 1 as well as for the CDM-based interviews in Study 2, due to some limitations during the research, participants for experiments using MATB simulation (Studies 3 and 4) were recruited from the public. For practical purposes, the flying task for these experiments was also simulated in a computer-based simulation instead of a flight simulator.

1.3.1 Covid-19 Impacts to the Research

The research was affected by the Covid-19 pandemic which occurred after completion of the brain and physiological study. The pandemic impacted the research in the sense that it was not possible to invite human participants for an in-person study due to safety, health, and legal concerns. The situation resulting from the pandemic has therefore forced a change away from the utilisation of a flight simulator or real professional pilots as previously planned. In addition, due to international travel restrictions, a planned research collaboration with a major aerospace organisation in the United States, focusing on physiological measurement in the context of sleep inertia, could not take place. Access was also restricted to the laboratory and the physiological devices that were to be used for the planned studies. Figure 1.1 below indicates the timeline of the planned and realised studies. As shown in the figure, studies conducted during the pandemic (CDM, Acceptance, and Prospective prediction studies) provided context to understanding studies reported in later chapters (Physiological studies). As such, a decision was made to bring them (CDM and Acceptance studies) forward within the thesis to ground the empirical work.

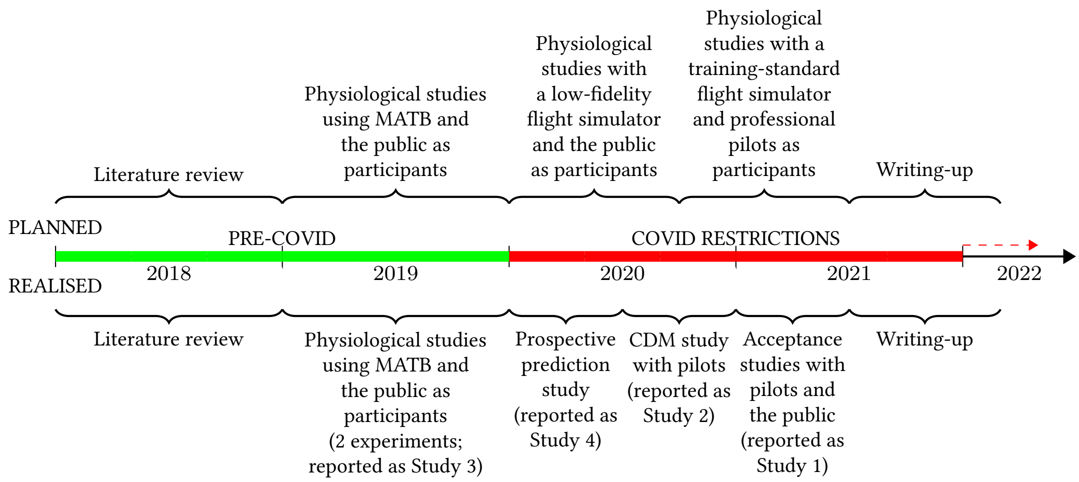


Figure 1.1: Timeline of planned and realised studies for the thesis.

These planned works would have fitted within the original thesis narrative through contributions to method development, expansion of analytical skills, and further understanding of the feasibility of MWL assessment technology in real aviation context. The planned works would have resulted in more robust conclusions regarding the feasibility of physiological methods, particularly fNIRS, in measuring MWL in a real work environment. This would have been useful in developing MWL assessment methods to be implemented in real aviation context. The planned works would have also contributed to the expansion of analytical techniques of real-time physiological data. In the context of this research, this would have been important for providing real-time MWL feedback to pilots.

With such restrictions, we developed an alternative research method to mitigate for any work that was prevented by the pandemic. The most feasible way of continuing our research during the pandemic was by leveraging online methods. Online methods made it possible to replicate conventional in-person methods such as lab experiments, field surveys, or interviews in a safer environment during pandemic restrictions. Considering this opportunity, we developed three online studies that would fit the objectives of the thesis.

To begin with, we decided to utilise the Critical Decision Methods (CDM) to explore cognitive demands on pilots when performing a crucial flight phase such as landing. This method was considered appropriate for supporting the narrative of the thesis because it focuses on the real experience of pilots as subject-matter experts. The results from this study therefore can support the use of MATB in the brain and physiological study in the sense of providing connection between the simulated task and the experienced task. This was important for improving the ecological validity of the brain and physiological study. In addition to the aforementioned support, the CDM study has provided a novel opportunity to explore the dynamics of pilots' MWL using a qualitative approach.

Furthermore, with the opportunity of conducting an online experiment, we decided to explore the use of the MATB task for prospective MWL prediction. Using an online experiment, this study aimed to test whether MWL in a simulated flying task can be predicted by merely observing the task; this provided additional narrative to the thesis regarding the feasibility to correctly predict the MATB task, including its subtasks, prospectively. This topic might contribute to the future expansion of MWL studies by including personal factors such as expectation to the task demand levels and its connection to the strategy when performing the actual task.

Finally, we decided to explore the perceptions of MWL assessment technology, held by the public and potential end users of MWL assessment. With the fact that real-time assessment of MWL is a work in progress, it is still less understood whether pilots and the public would accept the presence of MWL measuring technologies in the cockpit. The results from this study support the narrative of the thesis by providing dual insights about the attitudes towards the technologies from pilots and the public's perspectives; thus generally supports the endeavours of MWL assessment methods development. From this study, we also obtained insights about important factors in developing the design of the technology from the perspective of targeted users. This study was completed by conducting three online activities: (1) a survey and (2) a Focus

Group Discussion (FGD) for the pilots, and (3) an online experiment for the public. In summary, while the pandemic certainly impacted the intended trajectory of this work, the research that resulted from this thesis makes a meaningful contribution to the further understanding of MWL during a flying task and the way it can be measured; this includes the way pilots and the public perceive the emergence of the technology and its potential implementation in the future.

Chapter 2

Literature Review of Mental Workload and Its Potential Measurement Methods for Real-World Application

2.1 Chapter Overview

This chapter discusses the theoretical basis of mental workload in an attempt to find a practical framework to measure it in real-world applications. It also addresses the first aim of this research, which is 'to understand MWL assessment methods in the literature and opportunities for using physiological measures to estimate MWL in flying tasks.' To this end, this chapter presents a literature review on mental workload, defining the concept, underlying theories, practical frameworks for developing and evaluating it, and recent issues regarding real-world applications and operators' opinions about workload biosensors and technologies.

2.2 Definition of Mental Workload

Regarding the definition of mental workload, experts agree to disagree. To date, there is no single comprehensive definition that is widely accepted. However, there are several attributes that are likely to be mentioned when defining mental workload. The first attribute is mental processing capacity. It is argued that humans have a limited number of resources for processing information. The classic theory of 'the magic number seven plus or minus two' from Miller (1956) states that humans can temporarily store five to nine items of information in their working memory. As a heuristic, this remains popular to date despite that more recent studies have demonstrated different results. Gilchrist et al. (2008), for example, stated that young adults can store only three or four longer parts of information, such as idioms or short sentences. Meanwhile, Halford et al. (2007) proposed three to five chunks of information. Despite the debate regarding the actual number a human can store in their working memory, these studies show that our mental processing capacity is limited. This limitation aims either to conserve energy or to ease information recall in the future (Cowan, 2010). This limited capacity requires humans to select what information they need to be aware of to perform specific behaviour successfully. At this point, attention comes into the equation since it enables selection processes of information in three different manners: 'input selection', which directs processing to specific information; 'executive control', which manages ongoing tasks; and 'alerting', which will interrupt ongoing tasks to focus on new information (Remington & Loft, 2015).

The second attribute is task demand. Tasks, along with humans, seem to be the heart of human factors and ergonomics research. Much effort within the discipline aims to understand the way humans complete tasks as desired, particularly at work or in their daily lives. Hollnagel (2021, pp. 358) defines tasks as "any piece of work that has to be done, which is generally taken to mean the set of activities or functions that are needed to bring about a desired and intended outcome." Tasks and humans form an interaction. It can be simply expressed that tasks create demands on humans, either physical, mental, or both, and humans are required to meet these demands to complete the tasks. As human capacity is considered to be rigidly limited, tasks are the attribute that have to be 'engineered' first. The previously mentioned pilot's task priorities, i.e. aviate-navigate-communicate (Schutte & Trujillo, 1996), is an example of task engineering to help humans with limited capacity to achieve the goals of the task itself. With modern-day jobs creating more mental/cognitive demands (Sharples & Megaw, 2015), understanding tasks therefore plays an important role in mental workload studies. Tasks are the variable in an MWL study that are commonly manipulated by varying level of demand so that changes in mental workload can be observed and measured.

The third attribute is task performance. While a certain task may always create particular demands, the way a human operator may complete the task is a different story. One could assume that when task demands are high, operator performance will deteriorate. However, Sharples & Megaw (2015) explain that this is not always the case. Performance and task demands are not always negatively correlated as operators tend to monitor their performance and the feedback provided by the system e.g. from instruments, indicators, etc. This may alter their strategies to complete the task, perception of the task itself, and motivation in performing the task; thus, eventually altering their workload. Performance outcomes can also alter the consequent demands by changing the task. For example, a pilot missing their approach must perform other tasks to repeat approach procedures from the beginning. This almost certainly increases their workload.

The last attribute is subjective experience. Mental workload involves a degree of subjective assessment of task demands. The task or job may be identical, but different human operators will perceive experiences of MWL and situation appraisal differently. This pre-conscious process attempts to evaluate current levels of performance, arousal, and emotional responses, then adjusts effort allocation to mental or cognitive processing resources (Van Acker et al., 2018). Sharples & Megaw (2015) further explain that these psychological aspects are unavoidable as most tasks or jobs happen in work contexts, where external factors such as culture, support, and job type play essential role. Aside from these external factors, internal factors such as skill and motivation can also affect the way operators perceive workload and alter strategies to meet task demands, which consequently change experienced workload.

All or some of these four attributes may be included in a definition of mental workload, making definitions of the concept vary greatly. Since the debate around a definition continues, it is more useful to find common themes, for example, its end use; it is argued that the ultimate aim of developing an understanding of the mental workload concept is to "provide better, more comprehensive, and even more applicable, human-centered tools to promote both the efficiency and enjoyment of human work" (Hancock et al., 2021, pp. 204). Therefore, efforts must point in that direction instead of merely debating the definition. Understanding underlying theories of mental workload could be a plausible attempt to achieve that goal. The next section will discuss two most common theories: Limited Resource Theory by Kahneman (1973) and Multiple Resource Theory by Wickens (2008).

2.3 Underlying Theories of Mental Workload

2.3.1 Limited Resource Theory

The basic idea of Limited Resource Theory (LRT) (Kahneman, 1973) is that there is one central attentional resource for which all cognitive activities compete. As shown in Figure 2.1, when stimuli are present, a single-pool attentional resource will dynamically share attention capacity with each of them.

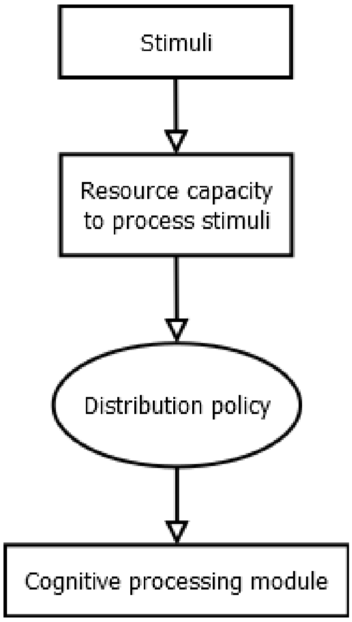


Figure 2.1: A simple graphical representation of Limited Resource Theory.

While mainly aimed at understanding attention, the theory can be extended further to understanding mental workload. Kahneman (1973) proposed that multiple tasks place demands on the central attentional resource, forcing the operator to choose a strategy and allocate attention capacity appropriately. Moreover, to understand capacity limits, there are three aspects of demand that need to be considered.

The first aspect is the 'demand level' imposed by the task. Tasks such as mental arithmetic, mental rehearsal, and timed activities (Kahneman, 1973), or any other tasks exercising working memory, are considered demanding because information in working memory degrades in a short period of time (Shiffrin & Schneider, 1977). Contrarily, data-limited tasks (Norman & Bobrow, 1975) or automated information processing activities need minimum attention to trigger a response, thus putting little demand on working memory and attentional resources.

The second aspect is the amount of attentional resource. This refers to how much capacity is available to an operator and what processes control it. The operator's ability to fulfil demands will be influenced by limited resource capacity. According to this theory, the limited resource capacity is flexible and elastic but not infinite (Navon & Gopher, 1979). The capacity available for processing information will 'inflate or deflate' depending on the level of arousal, mood, motivation, and age (Pickup et al., 2005).

The third aspect involves the rules concerning the allocation of resources. This aspect is related to the way in which human operators apply strategies to allocate the limited resource capacity. Wickens et al. (2012), for example, stated that people appear to favour heuristic-based techniques that give acceptable performance with minimal effort. The demands imposed on resource are therefore affected by individual preferences to perceive acceptable levels of performance and effort. Task difficulty may also be relevant to the allocation strategy as it is associated with subjective cost of effort (Pickup et al., 2005).

Wickens et al. (2012) proposed a graphical representation describing the model of Limited Resource Theory, as shown in Figure 2.2.

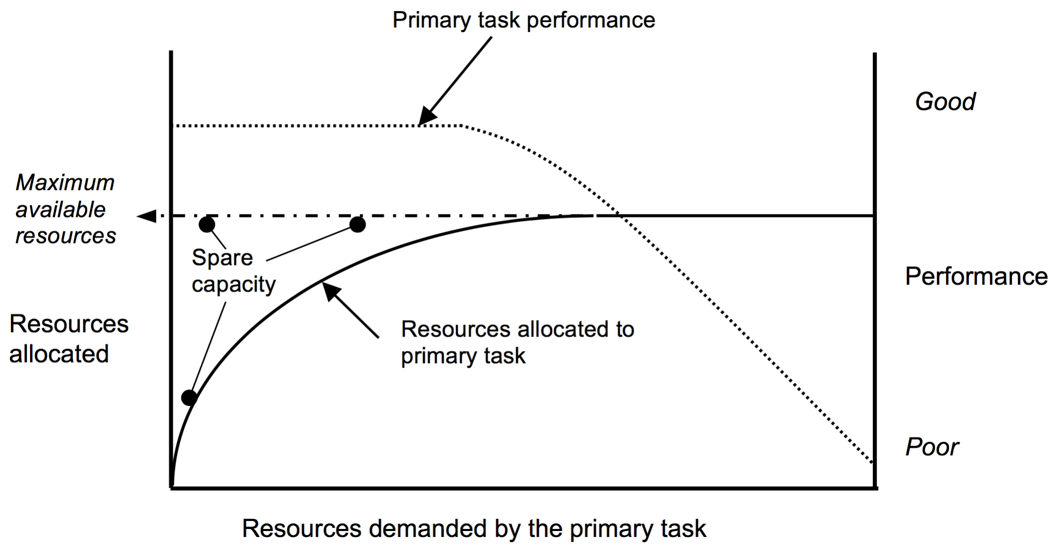


Figure 2.2: Limited resource model (Wickens et al., 2012).

From Figure 2.2, the link between the resources given to these various processes and the task demands imposed may therefore be conceptualised as mental workload. The left-hand vertical axis represents resources consumed by the task and the maximum number of resources available; the right-hand vertical axis represents performance in the primary task (dotted line). There is an area in the left of the figure where task performance is adequate since the available resources are greater than the job requires. Spare capacity or attentional resources can also be found in this area. The workload is inversely proportional to the quantity of spare resource capacity in this area, which reflects situations of low task demands. Moving to the right of the figure, there is an area where there are insufficient resources to fulfil the task demands as the resource limit has been reached. This is an area of high task demands and where workload and primary task performance are inversely related. As a result, in this area, a measure of primary task performance should be able to infer workload.

Nevertheless, Sharples & Megaw (2015) refined the graphical representation as multiple issues have been raised, as shown in Figure 2.3.

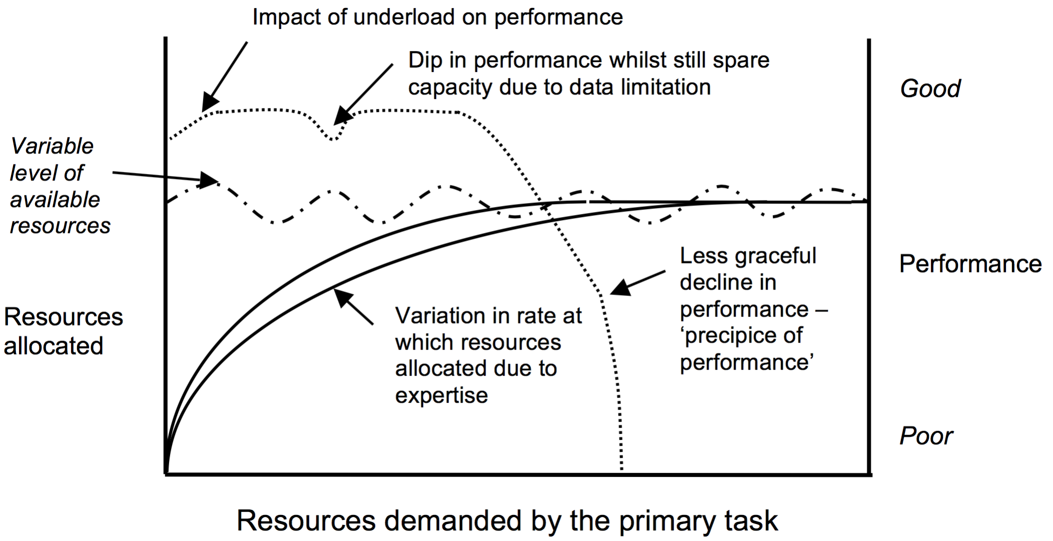


Figure 2.3: Modified limited resource model by Sharples & Megaw (2015).

A brief explanation of the issues is presented as follows.

1. External and internal influences: The model does not consider factors associated with the individual performing the task, such as stress or motivation. It is argued that psychological conditions may affect perceived workload and performance. These factors may alter the maximum level of available resources. Someone with low motivation, for example, might perceive an otherwise manageable task as highly demanding and thus performance may deteriorate. In this case, the maximum level of resources reduces.

2. Expertise and automaticity: It can be understood that the more experienced an operator is, the more easily a task can be completed. Sharples & Megaw (2015), citing Schneider & Shiffrin (1977), explains that this expertise is related to the ability to map elements of the tasks together. Consistent mapping is generally achieved once an individual has had plenty of practise with the task features. Expertise and automaticity may modify the consumption rate of the resources.

3. Underload: In addition to high-demand tasks, performance may also suffer when the demands are too low. This condition is known as 'underload' and

is more likely due to reduced maximum available attentional resources (Young & Stanton, 2002).

4. Graceful performance degradation: The model is criticised as the relationship among demand, effort, workload, and performance is not linear. Small increases in demand and effort may result in a 'precipice of performance' (Sharples & Megaw, 2015) and be detrimental to an individual's primary task performance.

5. Tasks, jobs, and work context: This issue represents the fact that in a real-world working environment, a job consists of a collection of tasks. The interactive nature of multiple activities with distinct temporal patterns, employing different sensory modalities, and imposing varying levels of difficulties, is not captured within the model.

6. Task switching: Related to the previous issue, task switching is inevitable under several conditions. A job consisting of multiple tasks may contain tasks that are not suitable to be performed in parallel. However, task switching is cognitively costly as it imposes unique demands and can have significant disruptive effect to the success of task completion in general (Hodgetts & Jones, 2006). The problem also comes from misperception of demands imposed by task switching, resulting in operators being hesitant to switch tasks and devote too much time on low-priority instead of high-priority tasks.

7. Data-limited versus resource-limited performance: As most of the tasks can be explained as a 'resource-limited process,' that is, as resources are invested, performance improves, yet often times the present volume of data available exceeds the system's capacity to handle it. At this point, the process turns into a data-limited process, where performance is no longer affected by the number of resources being invested (Norman & Bobrow, 1975). The model, however, does not represent these phenomena.

8. Single or multiple resources: Finally, it is strongly argued that the resources to deal with stimuli are not singular. They are more likely to be multiple, in terms of modalities to receive, process, code, and respond to the demands. At this point, the theory is extended to Multiple Resource Theory by Wickens (2008). The next section of this chapter will discuss the theory.

Together, these issues reflect the difference between strict laboratory experiments of attention and real-world experiences of workload. While studies in attention are primarily concerned with how contents of awareness of certain tasks are chosen (Remington & Loft, 2015), studies in mental workload seem to be broader since they involve various aspects such as the issues mentioned previously. This may suggest that the results of measuring attention and mental workload are related but not exactly synonymous.

2.3.2 Multiple Resource Theory

Contrary to LRT, the main preposition of Multiple Resource Theory (MRT) is, as the name suggests, that a human has multiple or separate information processing resources. The theory was not developed specifically for mental workload. However, it can be utilised to understand mental workload through understanding of the way humans perform multiple tasks, particularly in the context of time-sharing ability (Wickens, 2008).

Initially, there were three different information processing dichotomies, explaining the stages of cognitive processing, information encoding process, and modalities used for receiving information. The fourth element, the visual channel, was later added to the model. Figure 2.4 shows the 'cube' representing the model.

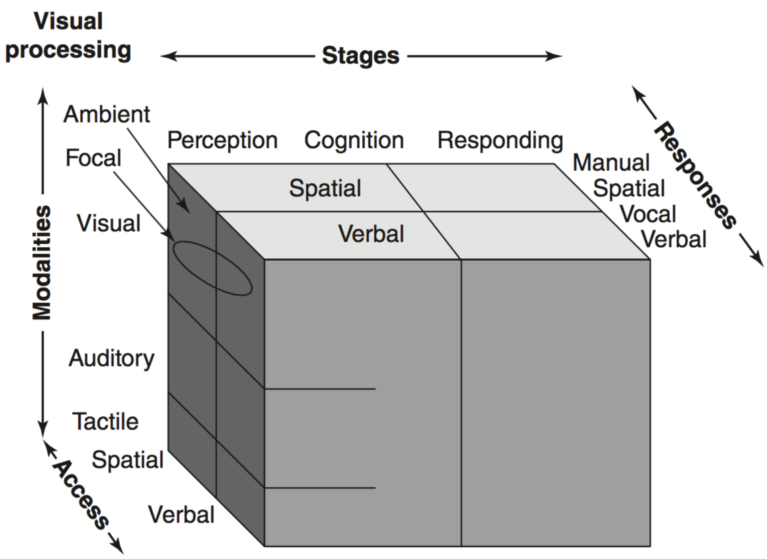


Figure 2.4: Multiple resource model by Wickens (2008).

According to this theory, there are three stages of information processing. Perceptual and cognitive activities are two of them. While perceptual activities are likely to be more simple (e.g. visual search) than cognitive activities involving complex tasks (e.g. decision-making), these activities exercise working memory and use a common resource to process information. Another aspect involves selection and response activities, such as speech production, which use its resource to process information. This resource separation affects performance in the sense that dual- task performance will not degrade as long as the two tasks are different in their processing stages (perceptual-cognitive vs response/selection). However, dual tasks with different stages but utilising the same resource (perceptual and cognitive) may interfere with each other. For example, having a conversation on the phone while driving may affect performance since both are using a common resource for perceptual-cognitive activities.

Regarding processing encoding, this theory describes a difference between analogue-spatial processing and categorical-symbolic processing, such as linguistic or verbal. The theory suggests that spatial and verbal processes, or codes, rely on different resources whether they are used in perception, cognition, or response stages of information processing. For example, it would be difficult to drive a car (spatial) while listening to an unfamiliar audiobook (verbal) as they have different processing codes.

Another essential dimension of this theory is perceptual modalities, or, the way information is received by our sensory 'devices.' The most common modalities for presenting a task are visual (the eyes) and auditory (the ears). This theory suggests that dividing attention between visual and auditory modalities will result in better performance than between two auditory or visual ones. This may be because tasks using the same modalities will drain the resource faster than tasks using different modalities. Checking a phone while driving would be more detrimental for driving performance than listening to the radio. However, apart from visual and auditory, tactile or 'touching' is recently considered as another perceptual resource channel. Stick-shakers in modern flight decks that warn pilots during stall conditions are an example of this part of the theory.

The last dimension of this theory is related to visual processing types: focal and ambient. It distinguishes between focal vision, which is involved in perceiving fine detail (e.g. reading a text), and ambient vision, which mainly involves peripheral vision (e.g. perceiving orientation). As a part of the common driving task, imagine that a driver keeping the car in the centre line of the road (ambient) while reading a road sign (focal). These tasks exercise different resources within the visual modality channel, and thus the tasks are likely to be executed successfully.

The MRT model has arguably survived the test of the time, since it can describe how different resource types may be considered whilst designing and evaluating task, while maintaining a human information processing viewpoint on mental workload. However, appreciating the context in which the workload occurs is also important. Mental workload should be understood in more contextual manner such as in a specific system. This view aims to retain the fundamental concept of mental workload yet at the same time emphasising the system context (Sharples, 2019).

2.4 Framework for Mental Workload Measurement

Aiming for making mental workload measurement more practical, Sharples & Megaw (2015) suggested a simple yet dynamic framework to understand measurement process and its implications. The idea of this framework is that operator mental workload, which is mainly subjective, is formed from the interaction of physical and cognitive task demands, performance feedback, and external and internal influences. Figure 2.5 shows the elements and their interaction.

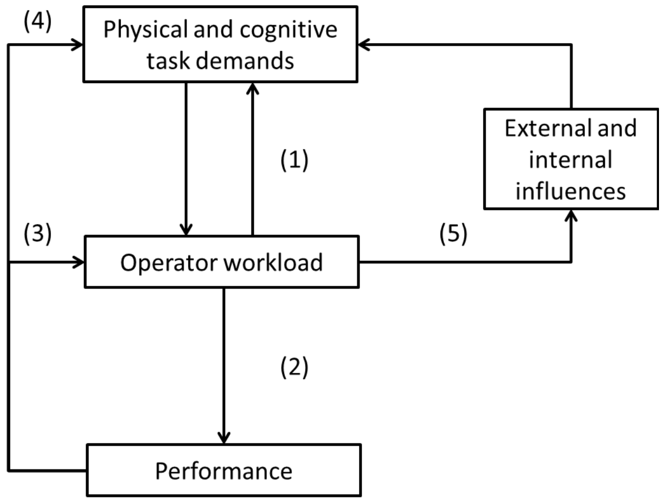


Figure 2.5: A dynamic framework for defining and measuring mental workload proposed by Sharples & Megaw (2015).

Operator workload is the effort or strain experienced by an operator when performing a task. This part is mainly subjective but can also be inferred from behavioural or physiological indices. The physical and cognitive tasks demands can be understood as ‘the job’ that must be performed by the operator. They specifically reflect both physical and cognitive elements of the task(s) that are most likely to be present together during completion of a task, and they will be affecting operator experience to the task. When performing a task, operators will be able to see the results of their job, that is, the performance feedback. This element is commonly measured using an objective metric, such as speed or number of errors. Finally, as mentioned earlier, external (e.g. company culture) and internal (e.g. motivation) influences are inevitable as much modern-day work happens in a social environment. This element, to some extent, plays an essential role in mental workload formation.

The interaction among these elements is also unique. It is not as simple as concluding that, for example, high workload will result in poor performance. A brief explanation of the interaction is presented below following the number shown in Figure 2.5.

1. Operator workload can be seen as a direct result of physical and cognitive demands. However, the framework suggests that external and internal influences must be considered since they may determine 'the magnitude' of the demands. For example, an identical task may yield different demands between novice and experienced operators.

2. Performance is assumed as a direct consequence of mental workload. However, the relationship is not necessarily linear. A highly demanding task may result in satisfactory performance, depending on the way operators maintain their performance. However, assessing how hard an operator works is considered challenging.

3. Operators can often become aware of their performance, either by using self-assessment or through displays or indicators. Performance feedback may then influence an operator's evaluation of their workload.

4. Not only affecting perceived workload, performance feedback may be able to change 'the magnitude' of the demands. Poor performance may produce other unexpected tasks to bring performance back to a desired level.

5. As pointed out in the first point, perceived mental workload may change task demands with the proxy of internal and external influences, such as selection of behavioural strategies.

The framework will be used as the guideline for studies in this PhD project because of its simplicity and practicality. Regarding the framework, a study from Maior et al. (2018) demonstrated the relationship between variables within this framework in assessing an operator's mental workload. In the study, the authors confirmed the relationship between physical and cognitive task demands and operator workload, i.e. that mental workload is subjectively perceived and objectively measured to increase as task demand increases. Consequently, performance might also be affected. Furthermore, when participants were made aware of their workload, their performance increased or decreased depending on the feedback type they were given. However, the study did not demonstrate direct relationship between performance outcomes and task demand changes.

This framework might be able to resolve the debates surrounding the definition of mental workload, and hence we proceed to the discussion of its measurement instead. Moreover, the framework might serve as a practical implementation of MRT in the sense of workload experience occurs continuously in the loop. Demands, influences, performance, and even mental workload itself can be considered as information that will be processed in the way that fits each cognitive resource. For example, information from air traffic control communications may be passed to the verbal/auditory modality channel for processing.

2.5 Measuring Mental Workload

There are many options regarding classification of mental workload measurement, and occasionally, they are complementary. Wickens et al. (2012), for instance, describes several types of measurement techniques, such as primary task performance, secondary task performance, behavioural indices, subjective measures, and neuroergonomics approaches. Meanwhile, Sharples & Megaw (2015) classifies the measurement into four categories, namely analytical, empirical, subjective, and psychophysiological techniques. The latter classification is preferable as it seems more structured in its organisation, providing better understanding of the variety of the measurement. The next section will discuss briefly about each category.

2.5.1 Analytic Techniques

The aim of analytic techniques is to understand mental workload by assessing and describing 'the job.' In other words, these techniques do not attempt to measure mental workload directly while a participant performs an actual task. The perspective of analysis can vary depending on the purpose of the measurement. Analysing the task seems intuitive, as previously mentioned, since mental workload is generated by the task. These techniques aim to evaluate mental workload along the timeline of task completion. In simpler words, workload during task performance is observed from beginning to end and then analysed to identify any time points corresponding to mental workload extremities. If unacceptably high workload is found, task refinement can be suggested. Pickup et al. (2010) provides a good example with the development of the Operational Demand Evaluation Checklist (ODEC) for understanding the demands of railway signalling works, by identifying quantifiable elements of rail signalling works and classifying them into high, medium, or low demand, based on subject-matter expert opinion.

With the task being used as the unit of analysis, understanding the people who perform the task in question are also relevant. Expert opinion can facilitate the understanding of potential workload experiences for certain tasks. The Critical Decision Method (CDM) is a popular technique to evaluate mental workload from a human operator's perspective. The method is "a semi-structured interview technique that uses cognitive probes to elicit information regarding expert decision-making" (Stanton et al., 2013, pp. 92). This technique is specially designed to understand cognitive demands during tasks or environments by eliciting information about cognitive functions such as decision-making, planning, and sense-making (Crandall et al., 2006). The output of this technique is a set of stories from subject-matter experts (SMEs) regarding the way they cognitively perform certain tasks. From the stories, we can classify demands and thus mental workload associated with the task performance.

It is also possible to compare workload data between two work systems, usually when one of those is still under development, in what is termed "comparability analysis." The aforementioned ODEC (Pickup et al., 2010) is an example of this. Data from an existing railway signalling works can be used for future development of a new railway signalling works.

2.5.2 Empirical Task Techniques

Compared to analytical techniques, empirical techniques attempt to understand mental workload by putting an operator into a real task. There are two types of technique under this category: primary and secondary task assessments. Primary task measures are based on direct assessment of variables related to the main task. For example, measuring an aircraft pilot's mental workload can be done by assessing their ability to manage the aeroplane in flight, using various relevant indicators such as altitude, speed, and horizontal position. From this example, the central idea of primary task measurement technique can be summarised as: the more difficult the task is (e.g. flying into a bad weather situation), the more the indicators will deviate from their desired target. Thus, mental workload can be inferred from these objective scores, i.e. wider deviation from the desired target may indicate higher workload. An issue arising from primary task techniques is that performance may not degrade following an increase in demands if the demands are still within the overall resource capacity of the operator (Young et al., 2015).

To resolve the issue, the secondary task technique comes in to play. The basic tenet of the secondary task technique is to provide a task that can compete with the primary task for the same attentional resources. The metrics from the secondary task can theoretically be used to determine the degree of MWL created by the primary task. The secondary task, in this context, can be employed as proxy for the spare resource capacity remaining from the primary task. Therefore, an increase in workload during the primary task will result in a decrease in spare capacity, making performance in the secondary task decrease. Using an example from a familiar driving situation, drivers must maintain the primary task of driving, for example, keeping within a lane, while being asked to set a satellite navigation system (satnav) if possible. The variance of accuracy of setting a satnav, for instance, can indicate mental workload required for the primary task. However, the main problem with secondary tasks is that experimenters cannot control the amount of attentional capacity invested in them, thus making them intrusive to the primary task under which the workload is being measured (Wickens et al., 2012).

One solution offered to tackle the problem is by using an embedded secondary task within the overall task environment (Raby & Wickens, 1994). The embedded task is naturally part of the overall task but less prioritised. An example of an embedded task is a check mark put by an air traffic controller when a plane has passed a certain navigational point. If the traffic in their sector increases, therefore increasing workload, they tend to delay or leave out the marking (Metzger & Parasuraman, 2005).

2.5.3 Subjective Techniques

As mentioned previously, operator workload is a highly subjective experience. Therefore, mental workload is often measured by asking how operators feel during or after performing a task. One of the most widely used tools to quantify workload is the NASA Task Load Index (NASA-TLX) developed by Hart & Staveland (1988). A copious number of researchers have employed NASA-TLX to capture operator workload level related to certain tasks, either as a primary measure for a dependent variable (Takae et al., 2010) or as a measure to determine level of mental workload in a pilot study (e.g. Fairclough et al., 2005; Hsu et al., 2015). The scale consists of six dimensions, namely mental demand, physical demand, temporal demand, performance, effort, and frustration level. Each dimension has a scale that ranging from “very low” (raw score “1”) to “very high” (raw score “20”; except for performance dimension that ranges from “perfect” to “failure”) that needs to be rated.

NASA-TLX is normally administered post-task, that is, when participants have completed the task. Notwithstanding the possibility to administer during a task, it seems inappropriate as the scale consist of six ‘questions’ to be answered, thus potentially jeopardising task continuity. The practical solution for this problem comes from the Instantaneous Self-Assessment of Workload instrument (ISA) (Brennan, 1992), that seems to fit the need to allow tracking of subjective workload changes during task completion. ISA, as the name suggests, provides immediate subjective evaluation of mental workload during a task and was originally developed for assessing air traffic controllers’ (ATC) workload. Its instantaneity makes this scale less-intrusive thus more capable for real-time assessment. The scale is using five-point rating scale to evaluate operator’s perceived workload (1 = low; 5 = high) and administered during a task with various intervals e.g. two minutes (Kirwan et al., 1997) or 45 seconds (Marinescu et al., 2018).

2.5.4 Psychophysiological Techniques

The logic behind the concept of psychophysiological measurement of workload is straightforward; an increase in workload will be followed by an increase in arousal, corresponding to activity within the autonomic nervous system (Sharples & Megaw, 2015). Development of these techniques is intended to allow continuous measurement of workload in the real workplace contexts. Before psychophysiological measures emerged, subjective techniques, specifically NASA-TLX, have become the most popular tools in workload studies. As previously mentioned, similar to other subjective techniques, NASA-TLX is innately retrospective with the administration of the tool undertaken post-task. It is also difficult to administer it during a task since the operator may become distracted. In operational settings, such as during flying an aircraft or driving a vehicle, administering subjective measurement tool tends to be impractical and can jeopardise safety.

Because of this, psychophysiological measures seem to be the rising star in workload research. The development of more advanced and more useful measuring devices has enabled the trend. There are several prevalent psychophysiological indices of workload that have been used to capture workload changes. Sharples & Megaw (2015) presented groups of psychophysiological techniques that are most widely used in workload studies, namely cardiac activity, brain activity, electrodermal activity, eye function, body fluid analysis, and muscle and movement analysis. Meanwhile, Kim et al. (2015) added photoplethysmographs (blood variables), electro-oculography, and skin temperature to portray the wide range of these measurement techniques. These techniques are used in various research contexts such as aviation (e.g. Ahlstrom et al., 2016), automobile driving(e.g. Foy et al., 2016), or brain-computer interfaces (e.g. Maior et al., 2015).

For the context of aviation in particular, a large number of workload studies using physiological techniques have been reported. Most of the studies applied electrocardiography (ECG) with its various variables to measure pilot workload (e.g. Dahlstrom & Nahlinder, 2009; Mansikka et al., 2016; Sauvet et al., 2009; Veltman & Gaillard, 1998). Newly developed wearable technology (e.g. Marinescu et al., 2018; Nixon & Charles, 2017) makes these techniques easier to administer in real and dynamic workplace settings such as in aviation, even in extreme aerobatic flying (Dahlstrom et al., 2011). Brain activity measures have also been utilised for workload studies particularly by using EEG (Dahlstrom et al., 2011; Wanyan et al., 2018; Wilson, 2002). Another technique for examining workload in aviation contexts is electro-oculography (EOG) which captures eyes activities through corneal and retinal potential (e.g. Dahlstrom et al., 2011; Veltman & Gaillard, 1998; Wanyan et al., 2018). Researchers have employed eye tracking devices (Di Nocera et al., 2007; Marinescu et al., 2018), facial thermography (Marinescu et al., 2018), and functional near-infrared spectroscopy (fNIRS) (Ayaz et al., 2012; Causse et al., 2017; Durantin et al., 2016; Marinescu et al., 2018; Takeuchi, 2000; Verdière et al., 2018), which are also quite popular in recent workload studies.

The commonality among workload studies is the use of multiple measures. Most of the studies used a subjective technique such as NASA-TLX to cross-check the data obtained from psychophysiological techniques. It is quite reasonable since workload measurement tends to be human-centred rather than task-centred, and thus subjective measures are likely to serve as a golden standard (Hart & Staveland, 1988). Most studies also claimed that they found good inter-correlation among measuring techniques whereas some of them did not. For instance, Wilson (2002) found some practical issues related to muscle artefacts when using EEG, making these techniques less useful during in-flight workload measurements. Another study investigated a range of physiological measures such as fNIRS, facial thermography, cardiac, and respiratory sensing, yet found the association between task demand changes correlated with and a subset of measures such as breathing rate and nose temperature (Argyle et al., 2021). This raises an issue to consider the practicality and reliability of measurement devices. For example, using EEG requires the experimenter to place several electrodes over the participant's head following the international 10-20 system. Similar to EEG, some fNIRS devices offer a full coverage head cap while some fNIRS sensors need to be put specifically in the forehead area. The use of EOG also requires placing electrodes on the participant's face specifically around (below and above) their eyes. Those techniques may not be a problem in lab-setting studies. Yet with current technologies, and considering the nature of pilot's job that mostly relies on visual and head function, the techniques are apparently still far from real-life implementation for pilot workload measurement.

2.6 Measuring Mental Workload in Real-Life Environment

Most mental workload studies are undertaken in lab-based environments, including simulators. However, the application of mental workload measurement is, indeed, valuable in operational situations. With the advancement of psychophysiological theory and techniques to measure mental workload, it has been suggested that research is needed to apply these measurements in real-world work settings (Midha et al., 2021). The main objective of real-time mental workload measurement is to provide feedback and alerts to an operator about their current workload. It is argued to be beneficial for assisting an operator in managing their task. For example, in a workload overload situation, feedback and alerts may be helpful to indicate their workload state and to identify possible actions to handle the situation (Maior et al., 2018).

Several studies in the past ten years have attempted to measure mental workload using various psychophysiological techniques in real workplace contexts, such as traffic control centres (Fallahi et al., 2016), driving (Lei et al., 2017; Sahaï et al., 2021; Schoedel et al., 2018), and electric bike riding (Boele-Vos et al., 2017). However, these studies did not provide real-time feedback and alerts about mental workload as in Maior et al. (2018)'s study in a lab setting. Even so, these studies may shed light on the plausibility of applying mental workload measurement, using psychophysiological methods in particular, in certain real-life work contexts.

As mentioned earlier, the study from Maior et al. (2018) indicated that workload feedback could be helpful for task management for some participants. Still, disagreement from the remaining participants also arose, noting that the feedback system could be stressful and could produce anxiety. From the study as well, it was revealed that feedback systems were prone to being ignored. With such opposing opinions, knowing what operators or subject-matter experts (SMEs) think about workload sensor technology and its application in real workplaces may contribute to the progression of mental workload research. Not only operators and SMEs, but the public's opinion is considerably important to understand as the application of such technologies may affect their lives. Imagine in the future, if pilots were mandated to wear workload biosensors to fly the aircraft, this would raise questions as to whether passengers would favour such policy. According to author's knowledge, this particular research topic is still sparse in the literature.

2.7 Chapter Summary

This chapter explored the theoretical basis of mental workload as a concept and practical framework to measure it. It is suggested that using a practical framework to measure workload is more favourable than merely debating the definition. From this exploration, it can be concluded that mental workload is shaped directly by task demands, both physical and cognitive, and operator performance. Internal and external influences, as well as performance feedback, may serve as proxies for adjusted task demands that eventually could alter the way operators perceive their mental workload. This framework will be used as the basis for the studies in this thesis. The next chapter will become the starting point for empirical chapters, describing the studies undertaken for this thesis and their results.

Chapter 3

Study 1: Professional and Public attitudes Towards the Application of Workload Sensors Technology in the Future Cockpit

3.1 Chapter Overview

This chapter presents a study examining attitudes from professionals and the public regarding a hypothetical scenario for the implementation of workload sensors technology in the future cockpit. It addresses the second research aim of this thesis. Three activities were carried out for this purpose. Activity 1 aimed to ask professional pilots their attitudes via an online questionnaire; meanwhile, Activity 2 aimed to deepen their response by inviting some of them for a focus group discussion (FGD) session. These activities were undertaken by using Technology Acceptance Model (TAM) theory as the framework for questionnaire development, interview guidance, and data analysis. Activity 3 aimed to investigate public preference towards the technology through an online experiment and using Willingness-to-Fly theory as its framework. This chapter presents insights about what professionals and the public think towards this particular issue.

3.2 Introduction

Technological changes are not new to pilots. In the past century, aviation has shown significant and fast changes, particularly regarding aircraft system technologies. The development of Traffic Collision and Avoidance System (TCAS), for example, was triggered by an air collision in the Grand Canyon in 1956 and finally, after years of perfecting the system, was formally implemented during the 1980s (Williamson & Spencer, 1989). This implementation requires pilots to understand how the technology works generally and how to respond to alerts accordingly (Williamson & Spencer, 1989). More recent technology such as Electronics Flight Bags (EFB) to replace paper-based charts and other essential information have also been implemented, particularly for general aviation (GA) operation (Winter et al., 2018a). Some of these technologies are mandatory (e.g. TCAS), requiring operators to install the device in the cockpit system and train their pilots in operating it; while some others, such as EFB, are apparently discretionary (Winter et al., 2018a), meaning that utilisation of the technology is decided by the operator, owner, or pilots. Whether it is mandatory or not, these technologies are the examples of changes that every pilot must accept and, perhaps, adopt throughout their careers.

Previous studies have been conducted to reveal pilots' acceptance towards particular aviation technologies. Richardson et al. (2019), for example, examined acceptance of F-16 fighter jet pilots towards the implementation of the Automatic Ground Collision Avoidance System (AGCAS), concluding that the way pilots perceive the usefulness of ACGAS may predict the usage of the technology. In other words, the more positive the attitudes are, the more likely pilots would accept the technology. Another example is the study from Fussell & Truong (2021), who examined the attitudes of student pilots towards the usage of virtual reality (VR) technology of flight training. The results suggested that the use of VR technology for dynamic learning such as in flight training can be predicted by, one of which, the way students perceived how useful and how easy-to-use the technology is.

However, these past exemplary studies investigated acceptance toward technologies that have been around for some time. This therefore raises a question of 'what would the pilots think or feel' towards imagined technology, that is, the technology that has not existed yet but aspire to emerge in the future, i.e. workload biosensors technology in the cockpit. Various attempts have been made to explore the efficacy of MWL measurements during a real task such as flying an aircraft (e.g. Wilson, 2002), indicating the possibility of future implementation in the cockpit. Therefore, examining pilots' attitudes towards this issue may benefit aviation in the sense of informing relevant stakeholders about how this future sensor technology should be implemented. Moreover, as the main interest of this chapter was actually professional pilots, we realised that public opinions matter too as they serve as one of the notable stakeholders in aviation. The public is a major 'consumer' of aviation, whose role is to generate monetary circulation for the industry. Any technological advancement in this industry is seemingly expected to achieve better operational, thus passengers safety. Therefore, they might also have an interest toward the implementation of the technology. This chapter attempted to address these issues and, for that purpose, has undertaken three activities. Specific research questions and hypotheses will be mentioned in corresponding activities.

3.3 Activity 1: Surveying Professional Pilots

As mentioned earlier in the overview, this chapter aims to address the research question about professional pilots and public attitudes towards the implementation of mental workload technology in the cockpit. Activity 1, and later Activity 2, attempted to examine the 'pilots' part of the question.

To develop insights about the issue in question, the Technology Acceptance Model (TAM) from Davis (1989) was used as the main approach. TAM is arguably the most influential, most tested, and best-operationalised approach to explain the acceptance of a technology because of cognitive factors (Schöpfel & Azeroual, 2021). TAM was originally developed for the context of information technology (IT; anything related to 'computers') during the mid-1980s, and heavily influenced by the theory of reasoned action (TRA) from (Ajzen, 1991). At that time, new IT technology had been made available for many work contexts, but resistance among managers and professionals existed (Davis, 1989; Davis et al., 1989). It is argued that to increase the use of IT, one should accept the technology first. These attitudes can be explored by questioning people about their future intentions to use the technology. Organisations would be able to manage the factors that influenced people's intentions to encourage acceptability and hence increase the use of IT technology if they knew what factors contributed to people's intentions (Davis, 1989).

The original model of TAM proposed behavioural intention (BI) as the most immediate antecedent of IT adoption, which is now commonly referred to as 'acceptance' (Davis, 1989) and occasionally the only measured outcome of interest (Chau & Hu, 2002). According to this initial model, one's attitudes (ATT) toward using technology has an impact on BI, and ATT has two predictors: perceived ease of use (PEU) and perceived usefulness (PU). Furthermore, PU is argued to have an independent effect on BI, while PEU affects PU. PU and PEU are arguably the core of this model, with PU defined by Davis (1989) as “the degree to which a person believes that using a particular system would enhance his or her job performance” (pp. 320), and PEU as “the degree to which a person believes that using a particular system would be free of effort” (pp. 320). Attitudes (ATT), however, is subject to debate among researchers in this area, considering it as a mediating variable with partial or full effect, or simply an irrelevant variable for ICT context thus needs to be discounted from the model (Kim et al., 2009). For this study, ATT was still considered as a part of the model since it is more proximate to the origin theory of TAM, which is TPB (Ajzen, 1991), and the evidence for removing of the ATT dimension mainly come from the ICT context (Kim et al., 2009). To our best knowledge, attitudes towards intention to use imagined technology, that is, those that have not existed yet, is still sparsely investigated. Figure 3.1 shows the relational diagram between these variables.

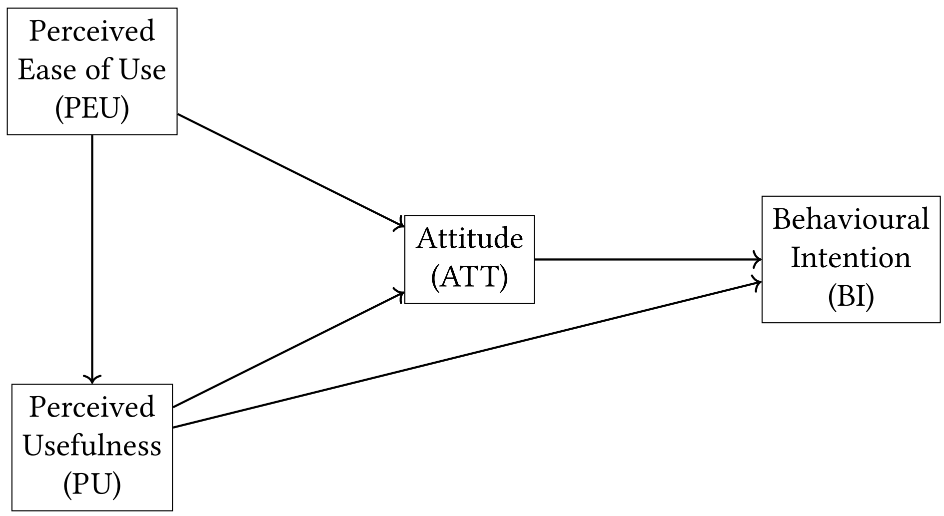


Figure 3.1: TAM as the framework for this activity.

Despite its origin for the context of IT technology, TAM has also been used in various contexts of technologies, such as online retail of financial services (McKechnie et al., 2006), mitigation system of driver distraction (Roberts et al., 2012), smartphone application for tourism purpose (Lin et al., 2020), online shopping platform (Ashraf et al., 2014), or even green technology and products (Anser et al., 2020), to name a few. This may suggest the model is flexible to be applied in various technological contexts.

Since its introduction by Davis (1989), TAM has been competing with other theories for a similar purpose. A model of PC utilisation (MPCU), for example, was developed to understand technology acceptance using six determinants, namely job fit, complexity, long-term consequences, affect towards use, social factors, and facilitating conditions (Thompson et al., 1991). The motivation model (MM) by Davis et al. (1992)

was applied to study ICT use and adoption using the notion of intrinsic and extrinsic motivation. The combination between models was also attempted. For example, theory of planned behaviour (TPB) from Ajzen (1991), which predicts intentions on an actual behaviour by looking at attitudes, subjective norms, and perceived behavioural control, was combined with TAM by Taylor & Todd (1995). This theory mainly decomposed beliefs structure in TPB model into factors that affect them. Moreover, Rogers (1995) proposed a theory that viewed technology acceptance from the perspective of innovation (innovation diffusion theory).

TAM itself was also extended by adding two determinants: social influences and cognitive instrumental processes. This model, known as extension of TAM or TAM2, believes that these two additional elements may affect one's acceptance of technology (Venkatesh & Davis, 2000). The most ambitious effort to comprehend technology acceptance was possibly coming from works of Venkatesh et al. (2003) which tried to bring all before-mentioned theories and models available together, creating a unified theory of acceptance and use of technology (UTAUT). This theory used main determinants, which are performance expectancy, effort expectancy, social influence, and facilitating conditions, along with some moderators such as age and gender. This theory was once again extended as UTAUT2 by taking the context of consumer use into account (Venkatesh et al., 2012). TAM obtained its most recent update in 2008 (TAM3) when Venkatesh & Bala (2008) proposed additional computer-related dimensions to be incorporated into the existing models; these dimensions were: computer self-efficacy, perception of external control, computer anxiety and computer playfulness.

With various theories and models available, we believe that the original form of TAM would fit as a starting framework of our study. The practical mission of TAM, as suggested by Davis (1989), is quite similar to the purpose of this study. Moreover, this study is an initial endeavour in identifying professionals' attitudes towards hypothetical technology, thus exploring the surface of this issue seems to be a good starting point. Still, TAM models have evolved into the most important model for predicting human behaviour toward probable technology adoption or rejection. Numerous studies have backed up the model's validity, emphasising its broad applicability to a variety of technologies. Therefore, to answer the research question in this study, a set of hypotheses were enlisted based on the original TAM model. Figure 3.2 shows a positionally adjusted diagram to better understand the relational annotation between variables in the original TAM model.

From the figure, the hypothesis statements for this activity are as follows:

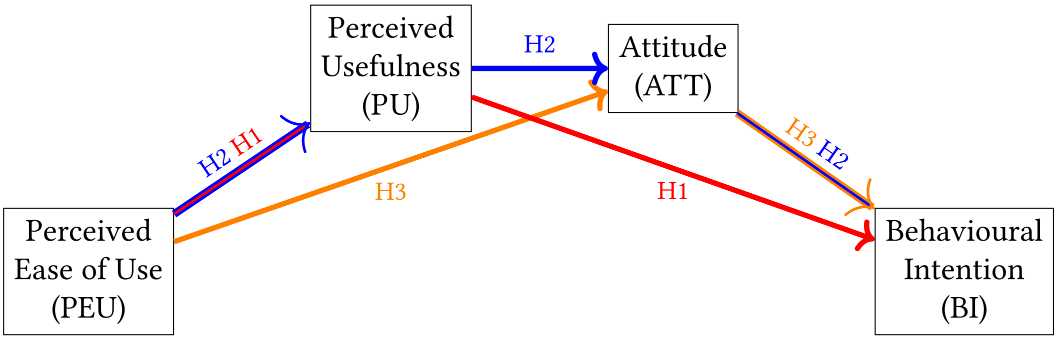


Figure 3.2: Hypotheses for Activity 1.

1. There will be positive correlation between PEU and BI through PU mediation (red pathway/H3.1.1).
2. There will be positive correlation between PEU and BI through PU and ATT mediation (blue pathway/H3.1.2).
3. There will be positive correlation between PEU and BI through ATT mediation (amber pathway/H3.1.3).

3.3.1 Methods

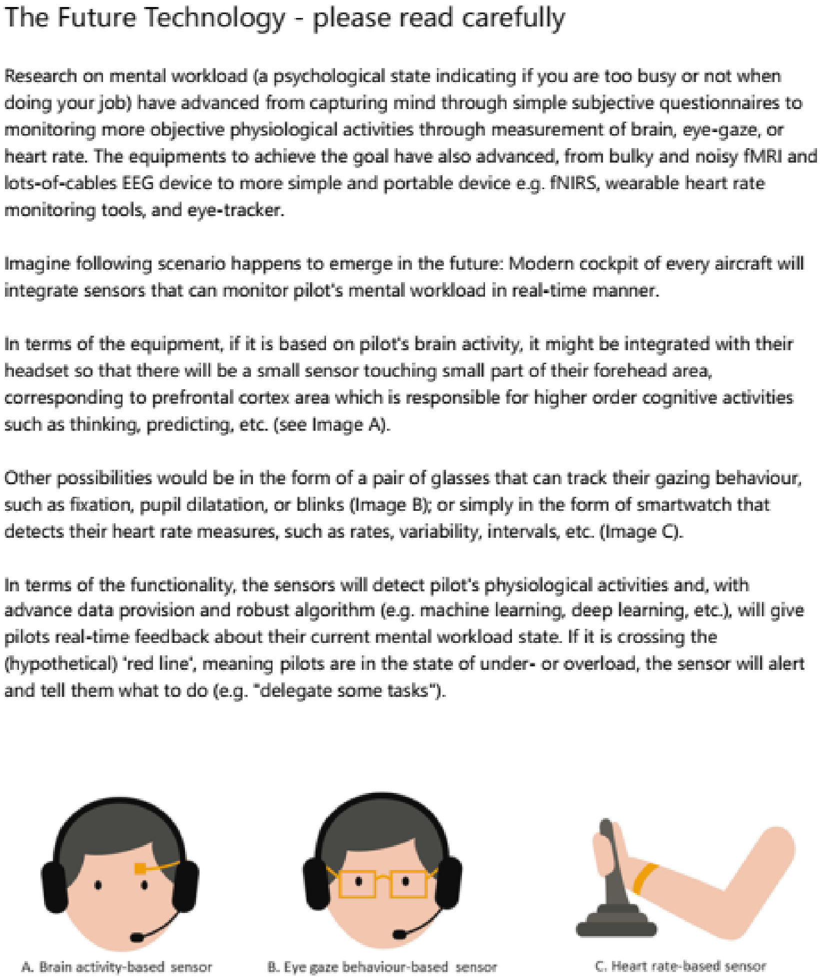
Study design and procedures. The design of this study was a survey with a close-ended questionnaire as the tool to gather pilots' attitudes. The TAM framework first suggested by Davis (1989) was utilised with adaptation for developing the questionnaire. PEU would act as an independent variable, BI as a dependent variable, and ATT and PU as mediating variables. We used survey methods for both theoretical and practical considerations. Since the purpose of our study was to gather information on attitudes of particular professional groups, and also was believed to be the first study exploring the issue, a descriptive and cross-sectional survey would be more appropriate (Kelley et al., 2003). For practical reasons, a survey was the most cost-effective and viable method for this purpose, especially amidst the period of restricted mobility (i.e. pandemic) where the group of professionals in question were severely affected. This restriction then rationalised the use of online survey for this study, as this technique seems more applicable in terms of its flexibility and ability to reach professionals. In this study, the questionnaires were produced using Microsoft Forms and the data was directly saved to the author's drive that was only accessible using credentials. The link to the questionnaire was controlled by the author and only made active during the period of data collection. Pilots who agreed to participate were asked to click the link and follow the instructions shown in the questionnaire.

Survey instrument. The questionnaire used Likert scale questions, with responses ranging from 1 ('strongly disagree') and 5 ('strongly agree'). In this study, the questionnaire was developed by the author from the original TAM theory and reviewed by experts, in this case, the author's supervisors. Apart from the main part of the questionnaire, we also gathered demographic information such as flight hours, gender, and current pilot rank. Table 3.1 shows the structure of the survey questionnaire; Appendix B.2 show the online version of this questionnaire.

Table 3.1: Structure of the questionnaire.

Factors	Code	Statements
Perceived Usefulness (PU)	PU1	Using workload sensors would make my flight safer.
	PU2	Using workload sensors would make me aware how busy I and my partner pilot are.
	PU3	I would trust what workload sensors alert me about my current workload.
	PU4	Workload sensor would be reliable, i.e. consistent in detecting workload dynamics.
Perceived Ease of Use (PEU)	PEU1	Using workload sensors would be easy.
	PEU2	Workload sensors would give clear and understandable feedback.
	PEU3	Learning workload sensors attributes (how to use, what the feedback means, what to do after the feedback, etc.) would be easy.
	PEU4	Using workload sensors would be comfortable.
	PEU5	Using workload sensors would not obstruct my flying tasks.
attitudes (ATT)	ATT1	I think it is a good idea to use sensors to monitor my workload.
	ATT2	I find it interesting to use sensors to monitor my workload.
	ATT3	Workload sensors technology is beneficial for the flight operation.
	ATT4	Workload sensors technology will have positive impact.
Behavioural Intention (BI)	BI1	I intend to use workload sensors once it is available in the cockpit.
	BI2	I intend to use workload sensors in every flight I am commanding.
	BI3	I intend to recommend my flying partner to use workload sensors.

The questionnaire also included a brief introduction of the hypothetical scenario regarding the implementation of MWL objective measurement devices in cockpit. This aimed to make the pilots aware of the new technology. The introduction sheet comprises the rationale behind the development of MWL sensor technology at the beginning. This leads to the introduction of possible forms of the sensors if they are about to be implemented in the cockpit. The description is also accompanied by the visual sketch of the sensors so that pilots can understand the scenario more comprehensively. Figure 3.3 shows the narrative of the scenario introduced to participants.



The Future Technology - please read carefully

Research on mental workload (a psychological state indicating if you are too busy or not when doing your job) have advanced from capturing mind through simple subjective questionnaires to monitoring more objective physiological activities through measurement of brain, eye-gaze, or heart rate. The equipments to achieve the goal have also advanced, from bulky and noisy fMRI and lots-of-cables EEG device to more simple and portable device e.g. fNIRS, wearable heart rate monitoring tools, and eye-tracker.

Imagine following scenario happens to emerge in the future: Modern cockpit of every aircraft will integrate sensors that can monitor pilot's mental workload in real-time manner.

In terms of the equipment, if it is based on pilot's brain activity, it might be integrated with their headset so that there will be a small sensor touching small part of their forehead area, corresponding to prefrontal cortex area which is responsible for higher order cognitive activities such as thinking, predicting, etc. (see Image A).

Other possibilities would be in the form of a pair of glasses that can track their gazing behaviour, such as fixation, pupil dilatation, or blinks (Image B); or simply in the form of smartwatch that detects their heart rate measures, such as rates, variability, intervals, etc. (Image C).

In terms of the functionality, the sensors will detect pilot's physiological activities and, with advance data provision and robust algorithm (e.g. machine learning, deep learning, etc.), will give pilots real-time feedback about their current mental workload state. If it is crossing the (hypothetical) 'red line', meaning pilots are in the state of under- or overload, the sensor will alert and tell them what to do (e.g. "delegate some tasks").

Figure 3.3: The hypothetical scenario introduced to participants.

Sampling and participants. We applied non-probabilistic sampling methods to obtain samples, particularly using purposive sampling. These techniques were considered appropriate for our study because purposive sampling is commonly used when a diversified sample is required or the opinion of experts in a certain field is sought (Martínez-Mesa et al., 2016). According to Battaglia (2008), the principal goal of purposive sampling is to generate a sample that may be assumed to represent the population, by "applying expert knowledge of the population to select a sample of elements that represents a cross-section of the population manner" (pp. 2). For this study, we set a delimitation for the population, which was 'active professional pilots' as the main characteristic. This characteristic therefore excluded general aviation pilots, student pilots, and fresh-graduate pilots without any experience in an airline (jobless). The type of aircraft to which pilots were associated were not limited, thus common dichotomous classifications such military or civilian, fixed- or rotary-wings (helicopter), and transporter or fighter jet were not relevant. Participants were then recruited by using personal approach to a certain group of pilots from Study 1. They assisted the author to spread the questionnaire (in the form of its hyperlink) to wider pilot communities. Ethics approval for this study was granted from Faculty of Engineering Research Ethics Committee, the University of Nottingham.

Data analysis process. Reliability analysis was performed using 'psych' package (Revelle, 2022) in R studio to check whether individual items support the construct of the questionnaire. Meanwhile, with the original TAM model we used for this study, mediation analysis was used as the approach to test the hypotheses since this approach may be able to explain how or why an independent variable has an impact on the outcome variable. We decided to apply structural equation modelling (SEM) because of its ability to produce a more appropriate inference framework for mediation analyses (Gunzler et al., 2013). For this analysis, based on the model, PEU acted as an independent variable, while ATT and PU acted as mediators. BI was still acting as dependent variables. Data analysis was performed using 'mediation' package (Tingley et al., 2014) and 'lavaan' package (Rosseel, 2012) in R studio. Figure 3.4 shows the analyses' pathway, with the letters to annotate coefficients.

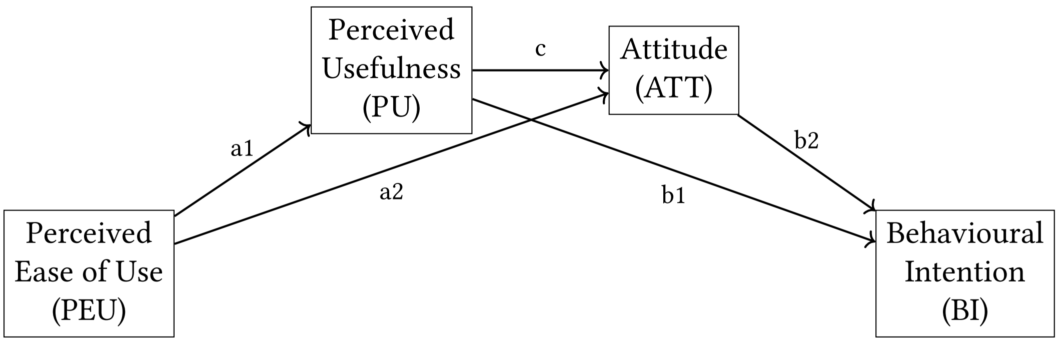


Figure 3.4: Pathway for the mediation process for Activity 1.

3.3.2 Results

Psychometric properties of the questionnaire. Internal reliability of the 16-item TAM-adapted questionnaire to gather pilots' attitudes towards application of workload sensors technology in future cockpit was investigated using Cronbach's alpha. Results indicated that the alpha for total scale was equal to 0.95. Examination of individual item statistics revealed that item deletion would not increase the reliability of the scale. Notwithstanding that the ideal alpha for making a good test is subject to expert debate and the purpose of the test, 0.7 was considered the minimum (Kline, 1999). Therefore, we believed that our questionnaire was sufficiently reliable.

We also investigated the reliability for each sub-scale that measured an individual dimension from TAM framework. The PU sub-scale indicated Cronbach's alpha of 0.83, PEU scored 0.79, alpha score for ATT was the highest with 0.90, and BI scored close to PU with alpha of 0.84. Based on the analysis, however, there were two items that would increase the reliability of the sub-scale if deleted. The items were the question of *"Workload sensors would give clear and understandable feedback"* in PEU sub-scale (coded as PEU_2) and *"Workload sensors technology will have positive impact"* in ATT sub-scale (coded as ATT_4). These two items would statistically increase the reliability to 0.81 and 0.92, respectively. Table 3.2 shows the complete analysis results for all items.

Table 3.2: Complete reliability analysis results of the questionnaire.

Factor	**Code**	**Correlation Coefficient**	**α if Item Deleted**	**Sub-scale α**
PU	PU_1	0.71	0.80	0.83
	PU_2	0.65	0.83	
	PU_3	0.79	0.78	
	PU_4	0.86	0.74	
PEU	PEU_1	0.67	0.75	0.79
	PEU_2	0.51	0.81	
	PEU_3	0.86	0.67	
	PEU_4	0.58	0.76	
	PEU_5	0.80	0.73	
ATT	ATT_1	0.94	0.82	0.90
	ATT_2	0.88	0.86	
	ATT_3	0.91	0.83	
	ATT_4	0.70	0.92	
BI	BI_1	0.89	0.73	0.84
	BI_2	0.79	0.84	
	BI_3	0.77	0.78	

These results suggested that the questionnaire was, in terms of psychometric, sufficiently good to obtain attitudes about the issue-in-question. Despite two less-reliable items, the reliability of the overall questionnaire was above the minimum standard of a 'good test' (Kline, 1999).

Descriptive statistics. Twenty professional pilots completed the TAM questionnaire (their average flying hours: M_{hours} = 3889.278, SD_{hours} = 1211.650). All participants were identified as male, with 10 pilots were ranked as 'captain' and the other 10 pilots were 'first officer' (F/O). The results show that 78.44 percent of the pilots' response was generally 'agree' with the implementation of workload sensors technology in the future cockpit (comprising the 'agree' and 'strongly agree' response); whereas five percent of them indicated disagreement (comprising the 'disagree' and 'strongly disagree' response). A fraction of 16.56 percent remained neutral or did not lean to either extreme. The results suggested that most pilots in our study seem to accept workload sensors technology when implemented in future cockpit. Figure 3.5 shows the proportion of participants' responses.

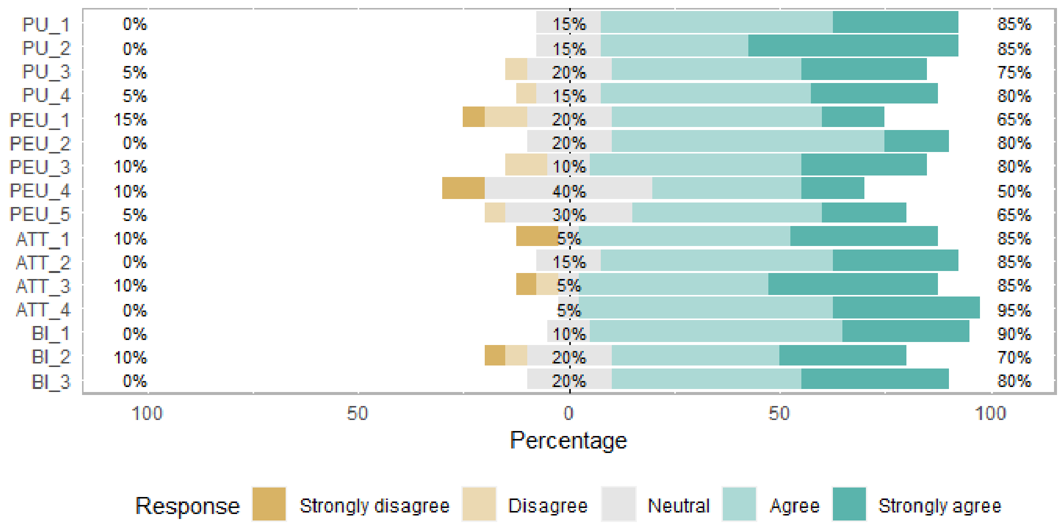


Figure 3.5: Summary of participants' responses.

Hypothesis testing. Before applying SEM for mediation analysis, we defined outcome, mediators and indirect effect models based on the pathway presented in Figure 3.4, as follows:

Outcome: BI = b1*PU + b2*ATT

Mediator 1: PU = a1*PEU

Mediator 2: ATT = a2*PEU + c*PU

Indirect effect 1: IE1 = a1*c*b2

Indirect effect 2: IE2 = a2*b2

Indirect effect 3: IE3 = a1*b1

The resulting model yields a good model fit (χ^2 = 83.832, df = 6, p = 0.00, CFI = 1.000, TLI = 1.073, RMSEA = 0.000) according to conventional criteria (Hooper et al., 2008).

SEM mediation analysis revealed that the direct relation between perceived usefulness (PU) and behavioural intention (BI) (i.e. path b1) is statistically significant and positive (β = 0.471, SE = 0.205, $p < 0.05$). Furthermore, the direct effect between attitude (ATT) and behavioural intention (BI) (i.e. path b2) is also positive and significant (β = 0.493, SE = 0.164, $p < 0.01$). The direct effect between perceived usefulness (PU) and perceived ease of use (PEU) (i.e. path a1) and between perceived usefulness (PU) and attitude (ATT) (i.e. path c) are also found to be positive and significant (β = 0.755, SE = 0.125, $p < 0.001$; β = 0.824, SE = 0.233, $p < 0.001$, respectively). However, the direct effect between perceived ease of use (PEU) and attitude (ATT) is not significant (β = 0.283, SE = 0.219, p = 0.197).

The indirect effect 1 (IE1) is positive and significant (IE1 = 0.306, SE = 0.143, $p < 0.05$), which indicates that perceived usefulness (PU) and attitude (ATT) mediates the association between perceived ease of use (PEU) and behavioural intention (BI). Moreover, the indirect effect 3 is also positive and significant (IE3 = 0.355, SE = 0.165, $p < 0.05$), indicating that perceived usefulness (PU) mediates the relationship between perceived ease of use (PEU) and behavioural intention (BI). Nevertheless, the indirect effect 2 is not significant (IE2 = 0.139, SE = 0.118, p = 0.236). This suggests that attitude (ATT) does not mediate the relationship between perceived ease of use (PEU) and behavioural intention (BI).

The results suggested that hypothesis H1 and H2 can be supported, while evidence for hypothesis H3 was insufficient. Since the TAM model we used did not indicate direct pathway between PEU and BI, direct effect analysis was not included in the analysis. The effect of each individual variable can also be seen from the analysis, as shown in Figure 3.6 (** indicates $p < 0.01$; * indicates $p < 0.05$).

3.3.3 Discussion

The results from this study indicate that the sampled pilots generally agreed with the implementation of workload sensors technology in the future cockpit. Their tendency to support this hypothetical scenario can be further explained by our data from the TAM-adapted questionnaire. From the data, pilots' acceptance of the technology can be started by looking at the way they perceive the easiness of using it (PEU variable). Conceptually, the ease of use can be defined as 'free of effort' when using any technology (Davis, 1989). However, following the model we used, PEU itself cannot directly cause the intention to actually use the technology. The perceptions of ease of use had to be mediated by the way pilots perceive its usefulness, or with addition to

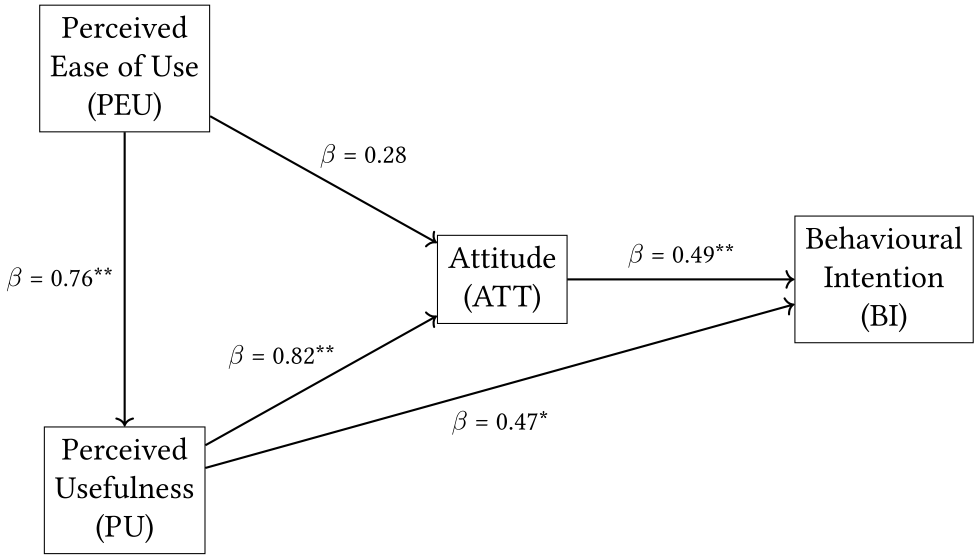


Figure 3.6: Relational effects between variables.

attitude mediation. We may infer from the results that perceived ease of use will lead pilots to perceive that the technology could enhance their performance (PU). This, consequently, will lead pilots to the intention of using workload sensors technology (hypothesis H1). The relationship between perceived ease of use and perceived usefulness was found to be strong, while the relationship between perceived usefulness and behavioural intention was moderate.

Similar to hypothesis H1, perceiving that the technology may work for them will lead to positive evaluation towards the technology first before forming the intention to use (hypothesis H2). Nevertheless, perceiving that workload sensors technology will be free of effort alone may not be sufficient to shape positive evaluation towards the technology, thus the indirect effect of this variable to predict the intention to use seems to be less supported. From the results, we may also explain that attitude plays an important role to shape pilots' acceptance towards workload sensors. Positive attitude may lead to behavioural intention of using workload sensors technology in the cockpit. As in the attitude variable, the way pilots perceive benefits of the sensors may directly and indirectly shape their intentions to use this hypothetical technology. In this study, the relationship between perceived usefulness and attitude was found to be strong, while the relationship between attitude and behavioural intention was moderate.

These results were similar to the results from Richardson et al. (2019). In their study, it was revealed that the ease of use has strong positive correlation with perceived usefulness in the context of integration of automatic ground collision specially developed for fighter jets. This perception, therefore, leads to behavioural intention to use the technology. However, similar to ours, the results from their study did not support the relationship between perception of ease of use and behavioural intention. Hence, our results were slightly different to the conceptual model of TAM in the sense that perceived ease of use may indirectly predict the intention to use a technology. One of the explanations for the disagreement was that, in our case, the technology-in-question was still hypothetical. Pilots might face difficulty to concretely translate the imagination of how easy the technology would be. Meanwhile, the degree to which the technology would theoretically benefit them might be easier to digest since it sounds a more abstract concept that does not require them to imagine actual use of the system. The statements asking them whether the sensors will give them clear and understandable feedback (PEU) might be more difficult to be evaluated than generic question about potential positive impact of the technology to the safety of a flight (PU).

The disagreement with the existing model and literature may also come from methodological issues. Only twenty pilots participated in this study, which can be considered a small number for survey studies. In a future study, a larger sample should be considered so that it would be possible to generate higher power in the results. With a considerable number of participants, confirmatory factor analysis (CFA) and omnidirectional relationship between variables could be properly assessed by SEM. Nevertheless, the results from Activity 1 have successfully informed us about the general attitudes that professionals hold with regards future implementation of the integration of workload sensors technology in the cockpit. Moreover, exploration regarding what forms this attitude may provide further insight. The effort could confirm, or contrarily, refute the results from this study, as presented in Activity 2. Identical aim to answer the same research question was set for Activity 2, yet with different methodology.

3.4 Activity 2: Exploring attitudes of Professional Pilots

The results from Activity 1 informed us that the surveyed pilots supported the integration of workload sensors technology in the future cockpit. Mediation analysis of the original TAM model also revealed how these attitudes happened. However, in-depth understanding of this attitude may inform us further about potentially hidden concerns towards such technology. Activity 2 aimed to answer the same research question as in Activity 1, but we applied a qualitative approach to explore more in-depth insight from professionals. The results of this activity are hypothesised to confirm our findings from previous activity.

3.4.1 Methods

Study design and procedures. We conducted online focus group discussion (FGD) via Microsoft Teams. Our rationale for choosing this technique was that FGD is considered a versatile method to discuss a particular design, prototype, or operational system from appropriate participants such as subject-matter experts (SMEs) (Stanton et al., 2013). A traditional way to perform FGD is by meeting participants in person. However, due to practicality and current mobility constraints, online FGD was chosen. It was made possible by the availability of more reliable online video conferencing apps and may be able to yield benefits such as increased satisfaction, flexibility of convenience, and comfort in discussing issues due to anonymity. This may solve several participant's issues such as time, mobility, health, and costs. Some contrary arguments, however, have also been addressed against online FGD, mainly regarding the lack of personal and non-verbal cues that can lead to misinterpretation, and selection bias caused by recruiting only computer-literate participants (Stanton et al., 2013). Since our target participants were professional pilots, computer-literacy issue appears to be irrelevant, and potential missing of non-verbal cues can be minimised by asking their concern to show themselves in-camera during interview.

The general procedure of FGD was applied as suggested by Stanton et al. (2013). We started by defining the aims and objectives, determining key discussion topics, assembling the focus group, administering a relevant demographic questionnaire, introducing the design concept (see Appendix B.1), and iteratively introducing topics to explore their attitudes. Table 3.3 shows details about these processes.

Table 3.3: FGD procedures.

Procedure	**Description**
Defining aims and objectives	This FGD aims to discuss the use of hypothetical use of workload sensors technology in cockpit using TAM framework.
Determine key discussion topics in logical order	Based on TAM framework, we determined four topics for FGD: PEU, PU, ATT, and BI.
Assembling a focus group	We invited three professional pilots per session using convenience sampling.

(continued on next page)

Table 3.3, continued

Procedure	Description
Administering demographic questions	We did not gather much demographic information as participants were invited from previous study (Activity 1) and this information has been previously collected.
Introducing design concept	We introduced hypothetical design of the technology in a narrative form with graphical illustration.
Introducing topics	At these points, we started the discussion as per order in key discussion topics. This was an iterative step until the moderator (the author) thought that the discussion can be moved to another topic.
Transcribing the data	*Explained later in the next section.*
Analysing the data	*Explained later in the next section.*

Sampling and participants. Six pilots participated in our study and were recruited in two distinct ways, mainly using a convenience sampling technique. The questionnaire in Activity 1 included an invitation to participate in this FGD study. If they were interested in participating, an inquiry asking for their email had to be completed. The author then contacted them using the email provided, with detailed explanation and attempt for matching their availability. We limited a maximum of three pilots for each session to ensure all participants' views or opinions were sufficiently represented with a smaller number of participants (Sharples & Cobb, 2015). Ethics approval for this study was granted from Faculty of Engineering Research Ethics Committee, the University of Nottingham. Table 3.4 shows the details of the participants.

Table 3.4: Participants for the focused group discussion.

Pilot code	Flight hours	Current rating	Current rank	Notes
P1	4500	CN-295	Captain	Medium military transport plane
P2	5672	Boeing 737-400	Captain	Modified for military transporter
P3	4800	Airbus A320	Captain	Civilian, ex-military
P4	5050	Airbus A320	Captain	Civilian
P5	6500	Airbus A320	Captain	Civilian, ex-military
P6	3500	Lockheed Martin F-16	Captain	Multi-role jet fighter

Data analysis process. Recordings of the interview sessions were automatically stored in Microsoft Stream and accessible only by using the author's credentials. The first step for analysing the data was producing transcription from the interview. We transcribed the interview using 'intelligent verbatim transcription' approach, i.e. including all participants' quotes but excluding excessive repetitions or word fillers such as 'um' or 'ee' (McMullin, 2021). We applied two ways to obtain transcription. If the interview was conducted using English, Microsoft Teams automatically produced a transcription; if the interview was using Bahasa (Indonesian language), transcription was performed by a hired transcriber to assure objectivity. Translation to English was then manually performed by the author. After getting the raw transcription, the author reviewed and performed correction if needed.

The next step for data analysis was coding. Since this study was guided by the TAM framework, we applied deductive coding. In deductive coding, a coding frame is used to construct a predefined list of codes before coding the data (Linneberg & Korsgaard, 2019). This coding approach focuses on topics that are recognised to be essential in the literature, and is frequently associated with theory testing or theory refining (Rowley, 2012). As this study was theory-driven, a coding framework could be derived from the theoretical framework (Linneberg & Korsgaard, 2019). The coding was an iterative process to make sure relevant excerpts from the FGD session were put into the correct coding framework. The final step was analysing the coded data to obtain a conclusion. Directed content analysis was considered appropriate for this FGD study as it fits the criteria for this analysis suggested by Hsieh & Shannon (2005), which are having an existing theory, aiming for a structured approach for the analysis, and using deductive approach in coding the data. Moreover, obtaining quantification through counts or percentage of statements may also be helpful to develop the narrative later. NVIVO software version 12 from QSR International was utilised to assist in coding and analysing the data.

3.4.2 Results

Regarding the FGD interview results, they are presented based on variables from the TAM framework. We classified and compiled excerpts according to their tendency to either agree and disagree with the corresponding questions. This, according to our interpretation, may resemble the questionnaire in Activity 1, thus provide linear explanation of pilots' attitudes towards the implementation of workload sensors technology.

Perceived usefulness (PU). The main keyword for this variable was, as suggested by its definition mentioned earlier, whether this hypothetical technology would *im-*

prove their performance or job in flying an aircraft. From the FGD results, we have found two major factors related to the way pilots positively perceive the usefulness of this technology. The first factor was self-awareness in knowing their stress level and ability. Pilots seem to be unaware of their stress level or condition, thus resistant to share excessive tasks to their flying partner. Moreover, pilots have a tendency to handle everything by themselves, since an egocentric personality resulting from difficult training and tight competition might play roles in this tendency. Therefore, mental workload monitoring can help pilots to acknowledge their limits. In case of a single-pilot fighter jet, the necessity of a stress or workload monitoring device was considered essential to remind them and prevent overconfidence. The excerpts from Pilot 1 and Pilot 6 below shows the supporting evidence:

> "This sensor is good to know our stress level so that we can share our tasks. Some pilots think they can do everything, so they don't need to share their tasks [P1]."

> "It will increase safety in the sense of preventing overconfidence from pilots [P6]."

The second factor we found was that the technology can help decision-making. More specifically, this technology may inform pilots of their stress or load level and help them to decide what to do when trouble arises. This may also encourage them to discuss possible problem-solving with partners. Pilots consider that if this kind of technology materialises, pilots will be able to develop awareness about their psychological condition. Furthermore, the technology was imagined to have the ability to override pilots in certain tasks; thus other tasks can be shared to human pilots. In general, pilots have positive attitudes towards this technology. Pilot 6, for example, stated that:

> "With good training, pilots will develop awareness about prioritising something more essential and crucial to safety. So, this technology may inform and alert pilots what to prioritise, so they know what to do to maintain safety [P6]."

However, apart from positive views towards this technology, we also found some dissenting attitudes and can be grouped into four issues. The first issue was about reduced roles of human in cockpit. This will, according to some pilots, change the current paradigm of flying which place humans above technology. That is, the human pilot is the one who should mainly fly the aircraft. The technology, no matter how sophisticated it is, should be placed in the second layer. Humans are the main decision makers, while technology assists them in making decision. One pilot, furthermore, argued about this issue with recent accident of Boeing 737-Max in Indonesia and Ethiopia. The pilot stated that technology cannot be one-hundred percent accurate, and thus humans are still needed. When the technology takes human authority too much, humans would face difficulty to take control of the aircraft, particularly in an emergency. Pilot 1 and Pilot 2 statements, respectively, supports this notion:

> "If this technology is implemented, this will change those golden rules (of flying). We must be sure about the accuracy and effectiveness of the technology. It will change the way we fly, our mindset too. We fly, we are the decision maker. I would say that this technology will shift our role to the second layer, while the technology is the first one [P1]."

> "But no matter how advance a technology is, it still can make error. If you remember Max (737 Max) crash, it was claimed as the most advance (plane), but still (do some) errors. The pilot cannot take over because the system has done too much. So, we cannot feel too confident that technology can be a hundred percent accurate [P2]."

Furthermore, this issue leads to the second issue: the validity of the technology. Pilots were curious about the validity of the measurements; whether the sensors can truly detect mental workload without being confounded by extraneous, unrelated variables such as family problems. It was argued that human error is complex and involves various triggers. Some of these triggers may be latent. Pilots, according to one participant, can enter the cockpit confidently without any visible stress, but it can arise mid-flight as the result of a combination of events and these latent variables. One pilot also came with the idea of workload associated with team-work. In dual-pilot operation, pilots share their tasks. The question, therefore, how can the sensors take the dynamic of this team-work environment into account in estimating pilot's mental workload. Pilot 1, for example, gave a statement regarding this issue:

> "Meanwhile, the negative effect of it would be...maybe it cannot capture conditions under teamwork environment. So, the homework after this technology arises in the future is to design the teamwork pattern of crews, so it will match to what the sensor will provide [P1]."

The third issue was about trust, and this includes trust to the technology and to flying partners. This issue seems to be associated with the issue regarding validity. Pilots are resistant to be overtaken completely by computer system. In this case, if the sensors detect and interpret an 'overload' condition of a pilot and the control is automatically overridden by the computer, this could be problematic. The computer can initiate actions that are different to what pilots think or plan to do. At this point, pilots demands a certain degree of 'protection' from this, for example, by providing immediate cancellation or manual override by human pilots. Moreover, reduced trust towards the flying partner might also arise as their mental workload condition will be made explicit by the technology. Pilots will know how each other respond to certain events, and this will affect their trust. One pilot argued this will produce unhealthy working conditions since they will develop their own assessment of other pilots. For example, a pilot might refuse to work with a certain pilot in the next duty because he or she is considered unable to manage the workload properly. Another pilot added an argument about the contagious effect of knowing a partner's workload. If a junior pilot, for instance, knew that his/her captain was in stressful condition, his/her psychological condition would also be affected. Pilot 5 statement shows part of the evidence of for this issue:

> "In my opinion, this technology will yield distrust among crews because we know our partner's ability. I would say that these sensors will deteriorate the mental and psychological state of pilots, instead of helping them [P5]."

The fourth issue mentioned was a security issue. Even though this issue was minor in quantity and only arisen by Pilot 2, the security issue is interesting to discuss. Pilot 2, as a military pilot, was concerned about the potential of data misuse from the technology. With massive use over long period of time, the technology may create a pool of data about pilot's ability in managing workload. If this data come from military personnel, e.g. pilots, and thus this might become a secret asset of a nation. Misuse or leakage of this kind of data would lead to a national security issue.

Based on the results, we may represent identified factors associated with perceived usefulness (PU) variable in accepting mental workload sensor technology among pilots as shown in Figure 3.7. Factors that are potentially favourable for pilot's acceptance are shown on the right side of PU.

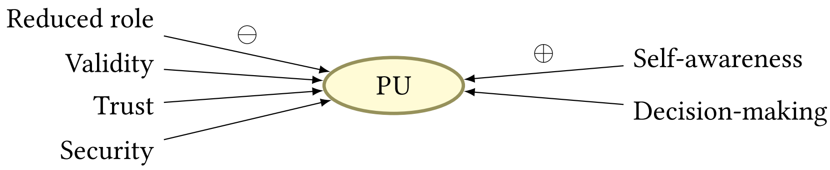


Figure 3.7: Identified factors associated with PU variable.

Perceived ease of use (PEU). Within the TAM framework, this variable attempts to measure acceptance of a technology from the perspective of whether the use of the technology will be *free of effort or less-effortful.* Participants apparently support the technology if it is easy to use and learn it, not limiting human authority, easy to understand the feedback, and comfortable to use. Pilots are always exposed to new technology in aviation, and it requires them to learn continuously. From the interview, pilots admitted that adopting new technology can be difficult at the beginning, but eventually, they will get used to it by continuously learning. They also believe that the technology is made to make their jobs easier and has been tested by their fellow pilots. Therefore, the acceptance of the technology may emerge from this sense of 'connectedness'. That is, it is made, tested, and used by humans. Pilot 1 and Pilot 3 stated, respectively:

> "For example, I flew a Fokker-27 that is less-advance than the C295. At the beginning I was forced to be familiar with CRM, then we finally are familiar with that. For me, we generally can follow the technology. We can adapt, even though (it was) tough at the beginning. Evolution of a technology needs trial and error so that it can end up as a robust technology [P1]."

> "But nothing is impossible. We can do something because we try. It would be learnable, I think, but maybe will take more time [P3]."

Nonetheless, from the interview, it was revealed that pilots wanted to maintain their authority to control the aircraft. Pilot 3, for example, explained that this technology must allow humans to override at anytime. This kind of technology, according to some of them, must provide pilots with a 'mode selector' feature. The feature allows pilots to choose the way the aircraft will 'behave' after getting input of mental workload data from pilots. For example, if the sensors detect 'overload' condition of pilots, this MWL technology must give at least three modes such as take control of the aircraft thoroughly (automatic mode), give suggestion on what pilots must do (advice mode), or pilots can just ignore information from the sensors (manual mode). Pilot 2 also communicated similar ideas with Pilot 2, believing that the basic principle of flying, i.e. humans, must be always the final decision maker. The statement of Pilot 2 represents this notion:

> "I prefer to be provided by a kind of mode selector. So, the final decision maker is us humans. No matter how perfect a technology is, human consideration is also important, to get rationale of why we have to do it manually, or automatic, or semi-automatic, for example. So, I prefer if we still act as a final decision maker [P2]."

Apart from the authority, the degree to which the feedback would be understandable seemed to matter for pilots in accepting the technology. Feedback in the form of auditory and visual modality were preferable as, according to pilots, that will be easy to understand. Moreover, pilots considered that physical representation of anything would help them to understand what is happening and what may have potentially caused it. Stress, for example, would be easier to understand if it is represented based on its physical underlying causes, such as blood pressure or heart rate. Presenting this representation on pilots' screen will assist pilots to decide what to do. Pilot 3 argued that:

> "As long as the feedback is in the form of audio and visual, I think it would not be difficult to understand [P3]."

Moreover, comfort was also considered essential by pilots, with the degree of comfort depending on how long the sensors had to be worn. Issues regarding comfort included three questions, specifically in what forms of modality the sensors will be, for how long pilots have to wear the sensors, and whether the sensors will be in the form of standalone device. Pilots favoured the sensors to be in the standalone form, meaning that the device on which the sensors will be attached is independent to other devices such as a headset. Smartwatch-based sensors was also preferable than brain and eye-gazing sensors, since attaching the sensors on the pilot's head would produce discomfort. However, this discomfort issue depends on the duration of use as well. If it is used for crucial phases only, such as take off and landing, that would be acceptable. However, Pilot 6, who is an F-16 fighter jet pilot, had a more flexible opinion regarding comfort. Pilot 6 argued that a fighter jet has been familiar with discomfort as they have to use many gears during flight. Adding this sensor's device, therefore, will not cause further discomfort. Pilot 4 statement, for example, can represent concern about comfort:

> "I would imagine using these sensors would be uncomfortable at the beginning. But it depends on the duration and period of use too. If it is used only at a certain phase of a flight, that would be acceptable [P4]."

Nevertheless, opposing statements disagreeing with the implementation of workload sensor technology were also found. The disagreement was mainly about the potential difficulty in using and learning the technology and discomfort when wearing the sensors. The difficulty in using and learning the technology might arise due to the nature of psychological aspects being measured. Pilots are trained in mostly, if not all, physical 'universe'. Instruments and control in an aircraft can be clearly seen and felt. For example, if a pilot turns an aircraft to the right, the result will be seen and felt immediately. Whereas, a psychological entity such as mental workload is considered 'too abstract' as it cannot be directly seen, thus difficult to be understood. One pilot also argued about the potential difficulty of learning for old pilots. Regarding ergonomics, eyeglasses-based sensors were less-favoured as this may obstruct pilot's scanning to the instruments in front. Moreover, mandatory face mask wearing during the pandemic situation might create fogs on the glasses. This device, if implemented, will also create a problem for pilots wearing correction glasses for the reasons of added weight to the head and reduced area of scanning. Some pilots concerned about the infrared light to detect brain activity if used for prolonged period of time. They worried about potential negative effects of the light to their flying abilities. Pilot 1's statement shows concern about difficulty in understanding the feedback, whilst Pilot 4's statement concerns discomfort, respectively:

> "Because we are not familiar with human behaviour such as doctor for example. We play with instrument, real thing. We command a plane to turn, then it will turn. So (psychological aspect of human) will be difficult to understand because it is too abstract [P1]."

> "But for the technology that interferes our eyes, our view will be limited, thus disturbing our scanning, our landing. We are required to see and touch the instruments [P4]."

Based on the results, we may represent the identified factors associated with perceived ease of use (PEU) variable in accepting mental workload sensor technology among pilots as shown in Figure 3.8. Factors that are potentially favourable for pilot's acceptance are shown on the right side of PEU.

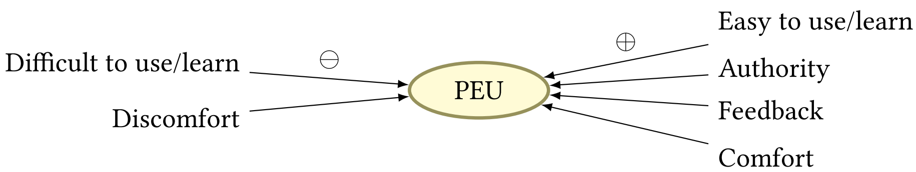


Figure 3.8: Identified factors associated with PEU variable.

attitudes (ATT). In general, we found that participants' attitudes towards this hypothetical technology were both positive and negative. The positive attitudes included opinions about potential improvement in aviation safety and minimising the probability of accidents, thus making people feeling safe when flying. Another statement revealed from the interview was about the possibility of this technology to be emerged in the future. Learning from previous technology in aviation, MWL sensor technology is not impossible. Previous technology such as auto-throttle, auto-trust, or autopilot in general were possibly considered 'impossible' in the past, yet they have become 'normal' today. Pilots also argued that this technology can be an opportunity for further research and development. Pilot 1 stated:

> "I am optimistic for this technology, and I am certain that God will help as long as we have a good intention on making this technology works. It can help aviation world, so we pilot will be safer and save many people. Hopefully, the research in this area advances so what becomes imagination for us today can be realised in the future [P1]."

Contrarily, some negative and pessimistic views were also found. Pilot 4, for example, talked about the technological paradox. According to Pilot 4, the impact of technology in reducing human cost of accident is not really significant. Pilot 4 argued that the human cost should have been able to be minimised with the advancement of technology. In fact, since 1970 the cause of accident mostly are humans. Generally, Pilot 4 believed that technology advancement does not always result in reduced human errors. Furthermore, Pilot 5 challenged the technology to be tested against time before gaining trust from the pilot community. Pilots tend to love aircraft that are easier to fly, and this technology might make things more complicated. For the moment, dual-pilot operation with proper training is sufficient for delivering a safe flight. Technology may assist, but after proving themselves to be reliable and 'trust-able'. Pilot 5 expressed this concern as follows:

> "This technology needs further research and must pass the test against time. It is similar to more recent issue of 'pilotless' plane, how can it build trust for passengers? It's the same case. The way we build trust is the most important for me, and it takes time. My opinion is based on my position as a multi crew pilot. We know our responsibilities, and the plane itself helps us with its advanced technology. In my opinion, pilots tend to prefer planes that are easier to fly. We are trained as a manager in cockpit. We have guides how to operate anything in the plane. So, it definitely takes considerably long time to gain trust among pilots [P5]."

Based on the results, we may represent positive and negative opinions shaping the attitudes (ATT) variable in accepting mental workload sensor technology among pilots as shown in Figure 3.9.

Figure 3.9: Positive and negative attitudes forming ATT variable.

Behavioural intention (BI). This variable theoretically served as the outcome within the TAM framework, that is, whether pilots have eventually used the technology once it is available in the cockpit. Since this technology was hypothetical, we asked pilots whether they *would use* if the technology were widely implemented. From the interview, we found that pilots support the usage intention of this technology if it has been tested and regulated. In other words, regulation of this technology was interpreted as assurance of the reliability in assisting pilots to make better decisions and improve safety. Pilots who felt assured with the technology would also recommend other pilots and their flying partner to use it as well. One pilot stated that regular usage of the technology seem to be unlikely. This pilot preferred it as a choice that can be used according to individual judgment. Pilot 2 and Pilot 4 said contrary statements respectively that covers these concerns:

> "For me, if it has been approved and regulated, comfortable, I will use it for sure. Particularly if it aims for safety and can help us to decide even better. And I will recommend to my FO if I were a commanding captain [P2]."

> "Personally, if this is a 'neutral' choice, I will use occasionally. I will not use from the beginning of the flight to the end. I use it when I want to know my mind and performance. For full-time use, I don't think so. For me, this is just for checking up. So, I can't always recommend, it's up to the pilots. I still believe that my workload has been reduced by the advanced system of the plane I'm currently flying. We just act as manager [P4]."

Based on the results, we may represent regular and occasional usage intention (BI) of mental workload sensor technology among pilots, as shown in Figure 3.10.

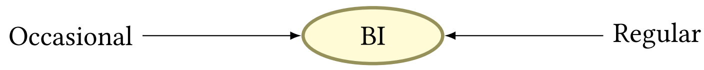


Figure 3.10: Usage intention (BI) of MWL sensors technology.

3.4.3 Discussion

This activity aimed to explore more deeply acceptance of pilots towards hypothetical technology of workload sensors in the future cockpit. Using FGD interview with six professional pilots, we identified factors that may influence their evaluation of the usefulness and the ease of use of the technology, thus affecting their general evaluation and eventually intention to use the technology. From a perceived usefulness perspective, factors that made pilots seem to accept the technology were its potential use for reminding them about their current psychological or stress condition (self-awareness), helping them in the decision-making process using information about themselves provided by the sensors, and general expectation about potential increase in safety. Pilots, as some of them said, tend to have a high-ego personality due to the prestige of the job. This sometimes would jeopardise safety in the sense of incorrect assessment of their ability. That is, they think that they are not in the state of 'stress' and can do all the tasks, yet in fact, they are highly stressful thus incapable of doing the tasks. The presence of this kind of technology is expected to moderate this personality, and thus they could see themselves more objectively. At the end, safety may improve because of more correct assessment of pilots' capability.

Along with these favourable attitudes, they also noted some potential disadvantages of the technology, creating a negative tone towards the implementation of this technology. These disadvantages involved a potentially reduced role of humans in the cockpit, questionable validity of the sensors, trust issues with partner pilots and technology itself, and security issues. The implementation of this technology in the future, according to some pilots, were potentially replacing humans' role in the cockpit. Humans are considered an essential part of the job for their abilities to produce desired actions and decision. They seemed to disagree if the technology totally takes over humans' authority in the cockpit. One of the reasons is possibly the traditional doctrine of flying, suggesting the human to fully control the aircraft when something has gone beyond expectation (P4). Along with this trust issue to the technology, measuring and explicitly showing their mental workload would lead to personal distrust among co-workers. Pilots may develop a stereotype of their flying partner based on previous information about their performance generated by the sensors, thus affecting attitudes towards them. This is considered 'bad' for team-work in the cockpit. Moreover, validity of the sensors were also questioned. To what extent the technology can be free of confounding variables, such as external factors, in measuring their mental workload might lead to disagreement to accept the technology. The security issue, particularly for military pilots, was also raised in the discussion, demanding reassurance of the way pilots' data will be treated.

The way pilots perceive this technology will be easy to use may shape their attitudes towards the implementation in the future. Four factors were concluded from the interview that seem to form positive attitudes to this technology, namely ease in using the sensors, ease in understanding the feedback, availability of mode selector, and comfort. With various technologies that have been applied in the cockpit, pilots are demanded to be a continuous learner. Regarding this technology, pilots would not mind learning once it is available in the cockpit, despite mentioning possible difficulty and longer time in familiarising the technology. This includes learning to understand the feedback provided by the sensors that is preferred in the form of auditory and visual modalities. Similar to the issue of reduced human role mentioned previously, pilots want to maintain their authority to control the aircraft. They suggested a 'mode selector' feature so that they can determine their level of authority before activation of the sensors. This is considered important since subjectivity may come into play in this matter. If pilots assess themselves 'fully capable', for example, they want the sensors to merely provide information, not taking over.

From the interview, the possible design of the device also became an issue. Capturing brain and physiological data requires the sensors to touch certain body parts, and this process is considered 'distracting' in the sense of creating discomfort and obstruction.

Pilots imagined about possible viewing limitation during landing, for instance, if they were mandated to wear eye-glasses. Pilots tend to feel comfortable if the sensors were designed in the form of a more familiar device, such as a smart-watch that can detect cardiac activity. The duration of which pilots are mandated to wear the sensors were also crucial. If the sensors must be worn for the entire flight, pilots would be most likely to refuse the technology. They preferred to wear it only in crucial phases such as take-off and landing.

These two variables, within the TAM framework, creates association with attitudes. FGD interview found both positive and negative evaluation to the implementation of the technology. The positive views mentioned by professionals were around the benefits for the aviation safety and appreciation to the potential pioneering works to realise this technology. However, some of them questioned the impact of the technology to reduce human cost of an accident, arguing that the reliability of humans cannot match the advancement of the aircraft system. In other words, as most contributing factors in aviation accidents in the past was the system or the aircraft itself, humans tend to be 'blamed' for more recent accidents. Furthermore, the interview concluded that most pilots would use the technology conditionally, if it has been tested and regulated. The rest of participants would use the technology as per their preference, for example, for checking their updated workload condition.

Within the TAM framework, it is argued that external variables affect perceived usefulness and perceived ease of use (Davis & Venkatesh, 1996). Since we explored acceptance of pilots towards hypothetical workload sensors technology in a more generic manner, we did not include possible external variables in Activity 1. However, Activity 2 results may inform us about potential external variables that construct pilots' evaluation to these two main variables in TAM. The external variables in TAM seem to vary depending on the context. Diop et al. (2019), for example, included variables that are relevant to adoption of variable message signs on the road, such as attitudes towards diversion, familiarity with road networks, and information quality. Scherer et al. (2019) mentioned different degree of explanation power of subjective norm, computer self-efficacy, and facilitating conditions to perceived usefulness and perceive ease of use in the context of teachers' technology. Therefore, results from this activity can be used to explore professional pilots' attitudes in more thoroughly manner. The relationship between Activity 1 and 2 are apparently reciprocal. We explored pilots' attitudes in generic terms (Activity 1), then attempted to explore further in Activity 2, from which the results can be used as external variables to explore pilots' attitudes using extended parsimonious TAM.

Activity 2 concluded that interviewed pilots are apparently willing to use workload sensors technology conditionally, with numerous external variables explained the way they perceived the usefulness and easiness of the technology. The next activity would explore attitudes of another important stakeholder in aviation: passengers. Workload sensors technology may not be directly relevant to passengers, but this implementation may become public concern since it will affect aviation safety in general, thus public safety specifically. Therefore, knowing public attitudes towards the implementation of workload sensors technology in cockpit is no less important.

3.5 Activity 3: Revealing Public Preference

From the FGD interview in Activity 2, one pilot mentioned the pioneering characteristics of workload sensor technology, saying that technology advancement in an aircraft have been mostly developed for increasing the reliability of the aircraft itself. Different to its previous developed systems in cockpit, this hypothetical technology focuses on the human aspect of a flight. This positive attitude might increase confidence and optimism towards creation, thus implementation of the technology in the future. At the same time, however, it is argued that new technology, particularly in the public and private sector, will always trigger public discourse (Podger, 2020). In the most recent example of the Covid-19 pandemic, the use of digital technologies have evidently helped to slow the pandemic by tracing, reporting, or educating the public, yet concerns about privacy, ethical, and centralisation of data are raised (Budd et al., 2020). In the context of transport, for example, members of the public seem to be concerned about reliability and potential job loss as the result of autonomous vehicles' presence in society (Hilgarter & Granig, 2020). In a democratic and open-information society, the public seemingly always want to know about new technologies, especially if they have potential impact to them.

Specifically in the aviation context, public concerns seem not to target the use of individual technologies. The public are apparently more concerned about the implementation of certain methods or products of aviation. For example, the implementation of pilotless aircraft does not appear to be attracting passengers. A study from MacSween-George (2003) revealed that only 10.5 percent of surveyed passengers were willing to fly in an unpiloted aircraft. This figure has recently changed, as a similar study in 2015 suggested that around 38 percent are willing to be a passenger in an autonomous airliner (Vance & Malik, 2015). Changes in the direction of research in aviation, particularly regarding unmanned aerial system and urban aerial mobility (UKRI, 2021), might contribute to the public attitude changes. Nonetheless, the study also suggested that passengers wanted to see that the aircraft can fly safely for a considerable amount of time beforehand. A similar case happened in the context of single-pilot airlines, where public concern about the potential danger of reducing flight deck crew from currently two pilots to a single pilot because of, part of the reasons, proneness to make mistake and inability to handle demanding tasks during a flight (Stewart & Harris, 2019). Apart from these technical and operational aspect, the public also may be interested about policies related to aviation, such as the use of biofuel (Filimonau et al., 2018) and aviation strategies to reduce the effect of climate change (Kantenbacher et al., 2018).

Studies attempting to understand public attitudes towards this potential technology in the cockpit are therefore essential, as this might become a public concern too. Activity 3 aimed to investigate public preferences regarding a hypothetical scenario in the future in which their pilots' mental workload are monitored lively during a flight using brain- or physiological-based measuring devices. In other words, we would like to know whether the public accepts the integration of such technology. To achieve this aim, we did not employ the TAM framework since it seems to be developed to understand technology acceptance from end-users perspective, that is, those who are actually using the technology directly. Passengers, in this context, are not using the technology directly but could be affected by its implementation. TAM, therefore, is no longer relevant for this context. Instead, we explored this issue from the perspective of consumers. Activity 1 and 2 have provided answers from 'professional pilots' part of Study 3 research question, whereas Activity 3 attempted to examine the 'public or consumers' part of the research question. Our hypothesis for this particular activity as follows:

1. Public willingness to fly a commercial aircraft would be different between these two hypothetical scenarios: when their pilots are flying while being monitored by workload sensors technology *versus* while not being monitored by the technology as in current situation (H3.3.1).

3.5.1 Methods

Experiment task and design. An online experiment was chosen for this activity. A between-subject design was applied for this experiment, with flying scenarios consisting of pilots being monitored and not being monitored acting as the independent variable, while the dependent variable was public willingness to fly. The task was a simple scenario to be read thoroughly by participants. The task itself consisted of two sections. The first section told participants about the possible emergence of workload sensor technology in the cockpit to measure or monitor pilots' mental workload. The detailed description about the device was also provided in this section. The second section was a hypothetical scenario in which participants imagined flying a commercial airline for a short-haul flight whose pilots either being monitored or not being monitored by workload sensor technology. They were then asked to evaluate their willingness to fly with such scenarios.

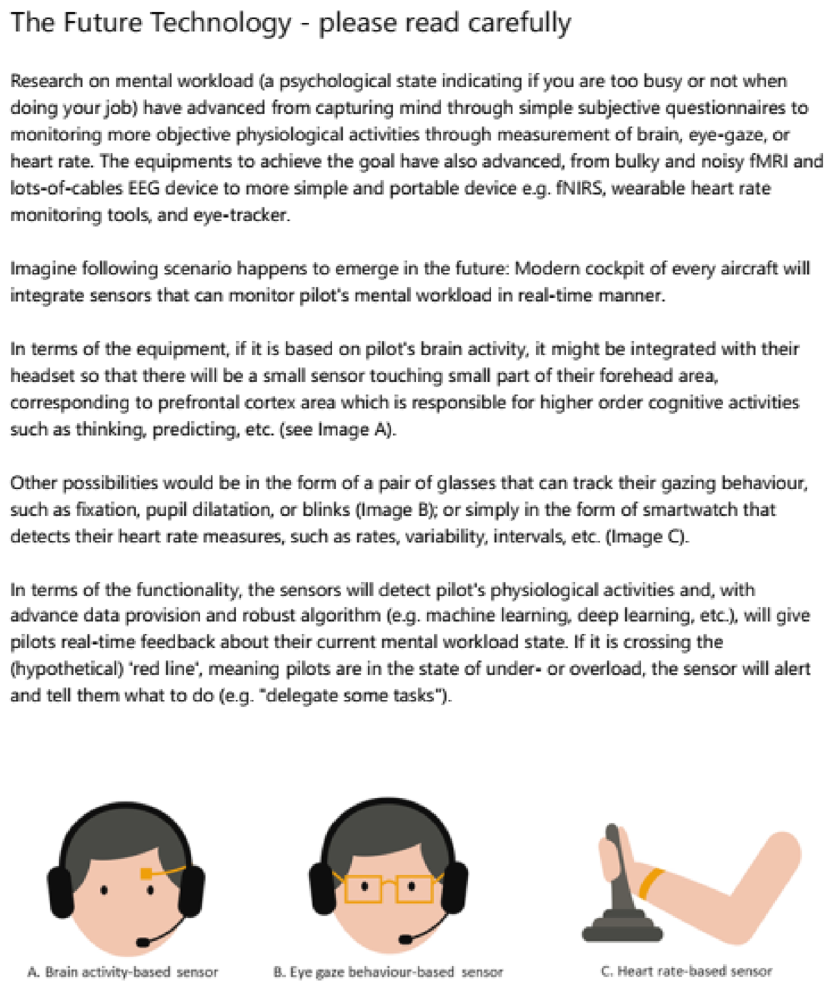
Participants. Fifty-seven members of the public (16 females) participated in this experiment. The average age of participants was 33.53 years old, with a standard deviation of 8.24 years old. Participants were recruited mostly from digital poster advertisement, distributed through various channels such as university emails, groups, or personal messaging services. If they agreed to participate, they followed a link provided in the poster to register their interest. Ethics approval for this study was granted from Faculty of Engineering Research Ethics Committee, the University of Nottingham.

Apparatus. This experiment was mainly using Microsoft Forms to put information about the study, informed consent, the scenarios, the questionnaires, and some demographic queries. There were two different versions of the form representing each scenario (monitored and not monitored) (see Appendix B.3). Heroku, a cloud service platform that enables a user to run an application entirely in the cloud, was utilised to randomly allocate participants. Regarding the questionnaire used in this study, it was adapted from the Willingness-to-Fly scale (Rice et al., 2020). Originally developed to identify issues regarding pilot and passenger adoption of new aviation technology, this scale has been used in different topics within aviation concerning passenger intention to fly in different technology-related scenarios e.g. autonomous commercial aeroplane (Rice et al., 2019) or intention to fly during the pandemic (Lamb et al., 2020). The scale was argued to be psychometrically satisfactory and versatile (Rice et al., 2020), even adaptations to other travel contexts such as maritime (Mehta et al., 2021) and auto-bus (Winter et al., 2018b) have been made possible. The scale comprises seven questions with Likert-style responses ranging from 'Strongly Disagree' to 'Strongly Agree', shown directly after presenting the scenario. The general score used for statistical analyses was obtained by averaging scores from individual questions. Table 3.5 shows the composition of the scale.

Table 3.5: Questions in Willingness-to-Fly scale (Rice et al., 2020)

No.	Question Wording
1.	I would be willing to fly in this situation.
2.	I would be comfortable flying in this situation.
3.	I would have no problem flying in this situation.
4.	I would be happy to fly in this situation.
5.	I would feel safe flying in this situation.
6.	I have no fear of flying in this situation.
7.	I feel confident flying in this situation.

The questionnaire also included a brief introduction of the hypothetical scenario regarding the implementation of MWL objective measurement devices in cockpit. This aimed to make the public aware of the new technology. Figure 3.11 shows the narrative of the scenario introduced to participants.


The Future Technology - please read carefully

Research on mental workload (a psychological state indicating if you are too busy or not when doing your job) have advanced from capturing mind through simple subjective questionnaires to monitoring more objective physiological activities through measurement of brain, eye-gaze, or heart rate. The equipments to achieve the goal have also advanced, from bulky and noisy fMRI and lots-of-cables EEG device to more simple and portable device e.g. fNIRS, wearable heart rate monitoring tools, and eye-tracker.

Imagine following scenario happens to emerge in the future: Modern cockpit of every aircraft will integrate sensors that can monitor pilot's mental workload in real-time manner.

In terms of the equipment, if it is based on pilot's brain activity, it might be integrated with their headset so that there will be a small sensor touching small part of their forehead area, corresponding to prefrontal cortex area which is responsible for higher order cognitive activities such as thinking, predicting, etc. (see Image A).

Other possibilities would be in the form of a pair of glasses that can track their gazing behaviour, such as fixation, pupil dilatation, or blinks (Image B); or simply in the form of smartwatch that detects their heart rate measures, such as rates, variability, intervals, etc. (Image C).

In terms of the functionality, the sensors will detect pilot's physiological activities and, with advance data provision and robust algorithm (e.g. machine learning, deep learning, etc.), will give pilots real-time feedback about their current mental workload state. If it is crossing the (hypothetical) 'red line', meaning pilots are in the state of under- or overload, the sensor will alert and tell them what to do (e.g. "delegate some tasks").

Figure 3.11: The hypothetical scenario introduced to participants.

Procedure. Participants agreeing to join the experiment followed a link provided in the digital advert that directed participants to a landing page in Heroku. Participants were then asked to follow another link containing a script commanding Heroku to randomly allocate participants to either versions of the form. Participants were instructed to read the information sheet, informed consent, and the scenario. These reading materials were followed by the willingness-to-fly questionnaire that needs to be completed by participants. Before ending the experiment, participants were asked to complete a few other demographic questions. Participants wishing to join a lottery of £10 Amazon vouchers for 10 winners were asked to leave their email address in a provided query.

3.5.2 Results

Descriptive statistics. As shown in Figure 3.12, there was a proportionate number of participants in both scenarios (sensor = 30, no sensor = 27). However, significant difference can be seen in gender (male = 41, female = 16) and flight frequency in a year (0 to 4 times = 34, 5 to 12 times = 19, and more than 12 times = 4). From these descriptive statistics, we found that there was no female participant who flies more than 12 times per year. The proportion between female and male participants was also considerably different. From these facts, we did not perform further analysis involving gender and frequency of flight category due to data insufficiency.

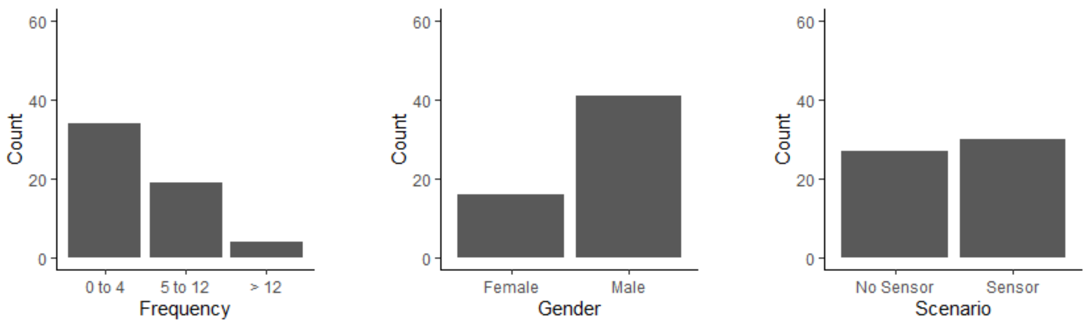


Figure 3.12: Descriptive statistics.

Testing hypothesis H3.3.1: sensor vs no sensor. With different sample sizes, we assumed unequal variance between two scenarios. Therefore, Welch's independent t-test as an alternative to ordinary Student's t-test was used (Welch, 1947; Ruxton, 2006). From the analysis, there was a significant difference in participants' willingness to fly score between the two scenarios ($t(53.55) = -3.47$, $p = 0.001$). The results suggest that members of the public might be willing to fly if their pilots' mental workload are monitored during a flight (M_{sensor} = 3.69, SD_{sensor} = 1.08) than otherwise (M_{sensor} = 2.66, SD_{sensor} = 1.15). Figure 3.13 shows the box plot representing the results. H1 of this experiment can be thus supported.

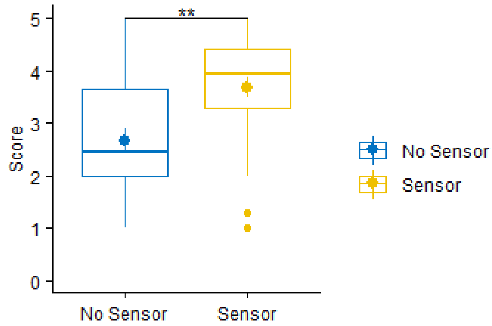


Figure 3.13: The results from hypothesis H1 testing. The dots indicate the mean from Willingness-to-Fly score with standard error of the mean. The difference is significant at p = 0.001.

3.5.3 Discussion

This experiment aimed to answer the 'public' part of Study 3 research question. More specifically, we would like to know if members of the public are willing to fly knowing their pilots are monitored using workload sensors technology. Willingness-to-fly score was used as an indicator for this tendency. From these experiments, it is argued that the public tends to have favourable attitudes towards the implementation of the technology, thus would be willing to fly if the pilots use the technology. The way we interpreted the results must be taken carefully. The two scenarios might be seen as contrary to each other yet, in fact, they both support the construct of interest. In other words, scoring low in 'no sensor' scenario has to be interpreted with positive attitudes towards the technology, since members of the public would tend to avoid the flight if the pilots are not monitored by workload sensors. The data showed that the difference in public attitudes, as measured by the willingness to fly score, between scenarios where pilots are monitored and not monitored by workload sensors technology were quite separate. Therefore, with that precautionary note, we might infer that members of the public, seem to support the implementation of workload sensors technology.

To our best knowledge, this experiment could be the first experiment addressing the issue of workload sensors technology and public perception to it. We may not be able to confirm our results with similar studies in the past, yet comparisons to the case of autonomous airliner (Vance & Malik, 2015) or single-pilot operation airlines (Stewart & Harris, 2019) may bring some insights. These two studies revealed public disagreement to the implementation of SPO or pilotless aircraft. Meanwhile, this experiment suggests that the public mostly agrees with MWL measurements of pilots during flying. Even though all of these before mentioned cases are innovative regarding the way the aircraft is operated, it appears that members of the public still need reassurance of their safety by the presence of pilots. Replacing one pilot, or even all pilots, may shape perception of declined safety among passengers. Contrary to these studies, the implementation of this hypothetical technology will not replace pilots but improve their reliability. Hence, the public may perceive this as an effort to increase safety. In our experiment, we did not specifically mention the benefits of the technology to participants e.g. increased safety. The scenario was only describing the technology from a technical and operational perspective, that is, how this technology works. Yet, the public may perceive that monitoring pilots' mental workload would be beneficial to improve safety in general thus, in a more specific manner, their safety.

3.6 Chapter Summary

This chapter presented three activities aiming to examine whether the hypothetical technology in measuring pilots' mental workload during flying operation could be accepted by both professional pilots and the public. From these three activities, the uniform attitudes supporting the implementation of this future technology could be seen. With deeper insight into professionals' opinions, several concerns or disagreements towards the technology were also found, along with greater strength of the attitudes that support the technology. The results might serve as preliminary exploration to this particular issue thus further endeavours are essentially encouraged. The next chapter utilises a qualitative approach to explore and identify what actually makes pilots cognitively 'busy' during a certain phase of a flight.

Chapter 4

Study 2: Critical Decision Method to Identify Sources of Pilot's Mental Workload during Landing

4.1 Chapter Overview

This chapter presents a study exploring how professional pilots perform real-world tasks during a particular phase of a flight. Chapter 4 also addresses the third research aim of this thesis, which is 'to explore cognitive processes of a real flying experiment from subject-matter experts (SMEs) and identify sources of task demands during a flying task'. It provides a description of their tasks during the landing phase of a real flight from the perspective of mental workload. Five professional pilots were interviewed using a Critical Decision Method (CDM) technique to recall a past flying experience. These interviews identified cognitive and physical demands imposed on pilots during a landing process, and how performance feedback and internal/external influences, along with the demands, interact with each other when landing an aircraft. This chapter provides insight about what actually makes a pilot cognitively 'busy', which can be used as a qualitative indicator of mental workload changes.

4.2 Introduction

Flying an aircraft is possibly one of the most challenging control tasks, since it needs a person with 'superior' cognitive and psychomotor performance to arguably produce safe flying performance (Hedge et al., 2000). Moreover, with modern aviation technology, the role does heavily rely on cognitive abilities of pilots. For example, Endsley & Bolstad (1994) described five critical abilities for safe performance of a flight:

1. Spatial, the capacity of a pilot to engage with aircraft systems through mental representation and spatial manipulation of objects that are important for navigation.

2. Attention, the ability to focus on important information during a demanding situation. Pilots have to be able to distribute their attention to several competing sources of information and tasks.

3. Memory, with two distinct features: working memory and long-term memory. Working memory is useful for comprehending situation and forecasting future events, as these tasks require the ability to gather and compare different sources of information for future prediction. Meanwhile, long-term memory is essential to store specific information during the flight so that it can reduce working memory load.

4. Perception, the ability of a pilot to process information quickly and stay alert to occasional signals to make quick decisions.

5. General cognitive functions, the ability to manage with high workload and to avoid certain issues under extreme pressure and environment during the flight.

These abilities contribute to situation awareness that is vital particularly in producing correct decision-making and task execution in a short period of time, which can be paramount for safety of the aircraft and, more importantly, passengers. From a mental workload perspective, these required cognitive functions may infer various sources of demands every pilot must handle during a flight. With such various sources of demands requiring outstanding ability in managing them, it may not be surprising that situation awareness has become the most contributing factor to commercial air transport accidents and incidents (Kharoufah et al., 2018). It can be said that insufficiency in situation awareness may lead to an accident or incident.

This study is therefore interested in capturing cognitive processes involved in a particular flying situation, such as landing. The landing phase was chosen since it tends to be more difficult thus generating higher workload compared to, for example, take-off, shown in objective (heart rate variability), subjective, and performance indices (Alaimo et al., 2020). Wilson (2002) noted that both take-off and landing are virtually crucial phases and produced significant changes in physiological data such as heart rate, EEG, and electrodermal indicator. However, landing is more visually demanding as this phase requires pilots to shift their gaze between inside (instruments) and outside environment (runway, obstacles, weather, etc.). Furthermore, in their analysis, Kharoufah et al. (2018) concluded that approach and landing are the phases of flight with most counts of observed accidents and incidents, around three times as much as take-off phase.

4.3 Methods

To capture cognitive processes of pilots during the landing phase, it was not possible for us to implement interview or observations *in-situ* due to safety concerns. Therefore, a technique called Critical Decision Method (CDM) was applied to understand cognitive demands during tasks or working environments by retrospectively eliciting information about cognitive function such as decision-making, planning, and sensemaking (Crandall et al., 2006). CDM is “a semi-structured interview technique that uses cognitive probes to elicit information regarding expert decision-making” (Stanton et al., 2013, pp. 92). Cognitive probes are the questions on each decision point that focus on particular aspects of cognitive processes and context behind the decisions made by the expert at the incident (Cattermole et al., 2016). For example, in the event of ’going around’ i.e. when the pilot decides to cancel landing, the questions to pilots might focus on cues used to initiate it, prior knowledge regarding the way it has to be done, the goal and expectation of this action, and possible options when facing similar situations. Cognitive probes are possibly the main key of a CDM interview.

CDM as an interviewing tool has been used for various contexts that exploit cognitive processes. For example, a study from Cattermole et al. (2016) utilises CDM to reveal naturalistic decision-making among experienced senior responders at traffic incidents. Moreover, in a similar context, CDM was used in a cognitive work analysis (CWA) to reveal issues regarding interagency collaboration, coordination, and interoperability of traffic incident management (Cattermole-Terzic & Horberry, 2020). CDM has also been used to create decision ladders to identify cognitive processes responsible for pilots’ responses during in-flight power-plant system malfunction (PSM) (Asmayawati & Nixon, 2020).

In this study, the CDM interview was open-ended and exploratory. We followed interview guidance from Crandall et al. (2006) to obtain in-depth understanding of pilots’ actions in each steps of the landing phase. At the beginning of the interviews, participants were asked to recall a specific landing experience of a flight they had been involved recently and, with assistance of the interviewer, completed a task decomposition diagram of the steps of the sampled landing experience. The task decomposition was essential to provide a timeline of events, which also served as a boundary and guidance for the interviews. In general, the structured approach (Wong, 2004) was used to analyse data from this interview, with some modifications to fit the aim of this study. The output of these interviews is a set of stories from professional pilots as subject-matter experts (SMEs) regarding the way they cognitively perform tasks during a landing phase. From the stories, their cognitive functions involved in completing the tasks are then classified according to a mental workload framework (Sharples & Megaw, 2015).

4.4 Participants

The CDM interviews were conducted with five subject-matters experts (SMEs) that included professional pilots from both military and civilian airlines. The pilots were from different airlines and had experience of flying different types of aircraft. The recruitment process used a snowball sampling method; one pilot provided a reference for other potential pilots to be interviewed, and so forth. Snowball sampling was considered relevant and more viable for this type of study, as the pilot community is generally limited in availability compared to the public. Institutional (sending a request to pilot community or airlines) and personal (companions, colleague's companion, etc) approach was applied to invite pilots to this study. Table 4.1 shows the details of the participants, with identities anonymised. Ethics approval for this study was granted from Faculty of Engineering Research Ethics Committee, the University of Nottingham.

Table 4.1: Participants for the CDM study.

Pilot code	Flight hours	Current rating	Current rank	Notes
001	6800	Airbus A330	First Officer (F/O)	Civilian
002	4136	CN-295	Captain/Instructor	Medium military transport plane
003	6500	Airbus A320	First Officer (F/O)	Civilian
004	4400	Airbus A330	First Officer (F/O)	Civilian
005	8730	Airbus A330	First Officer (F/O)	Civilian

4.5 Procedure

Participants were contacted via the interviewer's institutional email and provided with an information sheet and a consent form. If participants agreed to join the interview, a preliminary meeting was scheduled for each participant to briefly discuss the aim of the study and the CDM process, including reminding them to complete the consent form and scheduling the interview time. At this stage, participants were asked to prepare a sample landing experience to be discussed later at the actual interview. The flight could be any flight without specific restrictions regarding the time, the aircraft type, or their rank. The only requirement was that they must have acted as a Pilot Flying (PF) during a leg of the flight. After this briefing, an email was once again sent to participants with the link and the schedule of the interview.

On the actual CDM interview, participants were asked to recall a landing experience and to give an initial description of it. This aimed to share a common understanding between interviewer and interviewee. Then, participants together with the interviewer discussed in detail the steps that had been performed from the beginning to the end of a landing phase. The steps were then put into a task decomposition diagram. The process was iterative until the diagram was agreed by participants. The interviewer then provided participants with a set of questions to probe their actions and cognitive functions on each steps from the task decomposition diagram, using guidance adapted from Crandall et al. (2006). Each interview took between two and three hours and was audio-recorded. The interviews were conducted online by using Microsoft Teams.

After the interview, the recording was transcribed and analysed by the interviewer, assisted by NVIVO 12 software. Several participants were interviewed in mixed English and non-English language; thus, the translations were made available by the interviewer. A brief description, interview transcript, and task decomposition diagram for each participant was produced and sent back to them for correction. After several corrections, final version of these documents were used for further analysis. The interview recordings are automatically saved and stored in Microsoft Stream using the interviewer's university credentials.

4.6 Data Analysis Approach

Philosophically speaking, we assumed a constructivist approach as the interpretive framework in analysing the data, (Creswell & Poth, 2018) since it fitted our research purpose. This interpretive framework uses an inductive method of emergent ideas obtained from interview or observation (Creswell & Poth, 2018). Technically, once the data were ready to be analysed, we focused on the second level of the task composition diagram as these parts of the data served as our main interest. These parts of the diagram consist of main tasks that constituted the landing stage in a flight. We then treated them as 'workload point', that is, an event that was potentially contributes to mental workload changes during landing. For each workload point, we sought evidence from the interview transcript and matched them to the elements of the framework (from Sharples & Megaw, 2015), namely physical and cognitive demands, performance, and external/internal influences. This process was iterative to assure correctness of the matching. It may be possible that one excerpt matched more than one element of the framework, according to the researcher's interpretation.

4.7 Results

4.7.1 CDM Interview 1: Pilot 001

Description

Pilot 1, a senior first officer in a flag-carrier airlines, talked about the recent landing in Vietnam during the pandemic situation. It was a first-time landing in Ho Chi Minh City, and Pilot 1 was acting as Pilot Flying (PF) on the day. The aircraft, an Airbus A330-300, normally uses a passenger carrier configuration, yet on that day it was serving as a full cargo carrier due to Covid-19 restriction. It was a daytime flight.

Pilot 1 considered that the landing phase had started when passing 10,000 feet (3.05 km), while starting a procedure for approach (approach checklist), consisting of reducing speed, turning landing lights on, turning seat belts sign on, and cabin preparation for landing. Pilot 1 was asked by Ho Chi Minh City ATC to enter the aerodrome using RNAV STAR until the ATC gave them a radar vector. The instruction gave them a shortcut to the aerodrome so that the aircraft could get into the runway extension quicker as the traffic was not too crowded.

The runway extension procedure was started when Pilot 1 applied a 30-degree turn at 3000 feet (0.91 km) and was cleared for ILS runway 25R. The wind at that time was slightly gusty and made the aircraft bumpy. Thereafter, the ILS runway 25R procedure was started, around 12-15 miles from the end of the runway. Pilot 1 attempted to align the aircraft to the runway while reducing speed and applying flaps. When the glide slope signal was alive and captured, the aircraft was descending following its glide path. Pilot 1 subsequently asked PM (Pilot Monitoring) to put the gears down, continue to reduce speed, and apply full flaps. When the landing gears indicated ‘normal’ with three green lights illuminated, Pilot 1 initiated the landing checklist, and continued ILS approach visually. The weather was good with visibility was above 10 kilometres. The runway could be seen from far away, so Pilot 1 could decide to continue landing earlier.

At 500 feet (0.15 km), Pilot 1 decided to disconnect the autopilot, starting to fly the aircraft manually. The aircraft continued to land as Pilot 1 had visual of the runway. Pilot 1 successfully brought the aircraft to touchdown at runway 25R and continued rolling out up to the end of the runway. The aircraft vacated the runway by turning left, which directly put the aircraft into an active parallel runway, but no traffic appeared at the moment. The ATC cleared the aircraft to cross runway 25L, entered the taxiway, and parked the aircraft on the apron.

A task decomposition diagram of the steps for landing at Ho Chi Minh City was developed with Pilot 1, and can be seen at Figure 4.1 at the end of this section.

Workload points

Six mental workload points were identified and are discussed in the following tables (Table 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7).

Point 1: initiate approach. To start the approach, Pilot 1 used the altitude of 10,000 feet (3.05 km) as an indicator. The aircraft was descending from its cruising altitude of 41,000 feet (12.5 km) and had been cleared to 3000 feet (0.91 km). When the altimeter read 10,000 feet (3.05 km), the approach checklist was initiated by saying ‘approach checklist’ and this triggered PM to read out loud the items on the checklist. Pilot 1 responded with corresponding actions and confirmed to PM that the action has been done. First, the landing lights were turned to ‘on’ position. Second, PM asked about briefing for approach. As it had been done, Pilot 1 said ‘checked’. Third, seat belts sign was turned to ‘on’.

Table 4.2: CDM interview 1: point 1 – initiate approach.

Physical and cognitive demands	Performance	External and internal influences
ATC asked to descend to 3000 feet (0.91 km) from FL410. Altimeter 10,000 feet (3.05 km) was used as a clue for initiating approach checklist. PM read the checklist items while Pilot 1 responded with actions and confirmed.	Briefing had been done before descend.	Pilot 1 has memorised the checklist item as it has been done plenty of times.

Point 2: radar vector. In this stage, the pilot's goal was to bring the aircraft to the intercept point with a 30-degree bank angle. ATC gave changes of direction (heading) three times, thus Pilot 1 turned the heading knob as instructed by ATC and confirmed to PM that heading has been set. PM was responsible for the communication, but Pilot 1 was also listening to the radio, focusing on the heading change instruction. Because radar vector made the flight plan slightly changed, Pilot 1 attempted to estimate performance of the aircraft using 'raw data'. The conclusion made by Pilot 1 was that the aircraft would end up slightly higher as the rate of descent was decreasing as the result of reduced speed. Pilot 1 applied speed brake up to nearly 3000 feet (0.91 km) so that the aircraft would not be too high when intercepting.

Table 4.3: CDM interview 1: point 2 – radar vector.

Physical and cognitive demands	Performance	External and internal influences
Pilot 1 listened to radio communication between PM and ATC, and made heading changes as instruction and confirmed. Pilot 1 also performed rough estimation and calculation, concluded that the aircraft would be higher when intercepting, and thus applied a speed brake.	Rate of descend was higher after applying speed brake, from 1000 to 1500-2000 feet per minute.	According to Pilot 1's experience on past flights, if ATC directs the aircraft directly to near the runway, the altitude will be higher.

Point 3: intercept runway extension. After getting clearance from ATC, the aircraft was turning with a 30-degree bank angle to intercept runway extension. The

speed was reduced automatically from 250 to 210 knots. Pilot 1 check the route changes that had been done by PM due to radar vector.

Table 4.4: CDM interview 1: point 3 – intercept runway extension.

Physical and cognitive demands	Performance	External and internal influences
The speed was automatically reduced, so Pilot 1 only monitored and made sure that the speed was at the limit for 'green dots'. Pilot 1 received confirmation of route change from PM and checked it on FMCG.	The aircraft was slightly too high as predicted, and the route change was displayed on the monitor.	Reducing from 250 to 210 is a normal procedure for the aircraft. Pilot 1 also had MSA as reference for surrounding terrain.

Point 4: ILS procedure. ILS procedure was started around 12-15 miles from the end of the runway. The aircraft automatically attempted to align the aircraft to the runway while Pilot 1 was monitoring speed reduction and applying flaps, from 'up' to flaps 1 and then flaps 2. Pilot 1 told the command for the flaps and PM executed it and confirmed to Pilot 1. Pilot 1 also monitored radio communication and frequency changes. When the glide slope signal was captured and alive, the aircraft was descending following the glide path, and Pilot 1 commanded to lower landing gears and applied flaps to 'full'. The speed was still reducing automatically following computer plan. Pilot 1 monitored the green light indicator to check if the gears have been down and locked. Pilot 1 then initiated 'landing checklist', consisting of (1) autobrake, which had been set; (2) auto thrust, which confirmed by PF that the setting was 'speed' for automated speed using FMCG plan; and (3) ECAM memo, which had fitted landing configuration and no missed item.

Table 4.5: CDM interview 1: point 4 – ILS procedure.

Physical and cognitive demands	Performance	External and internal influences
Pilot 1 asked PM to lower flaps from 'up' position to 'full' one-by-one according to the speed. Pilot 1 also listened to radio communication and monitor frequency changes, as done by PM. Pilot 1 monitored glide slope and asked PM to lower landing gears. Landing checklist was requested after all of these actions had been completed, with Pilot 1 set autobrake and confirmed auto thrust and ECAM memo status.	All technical actions feedback were shown on the display.	Pilot 1 checked what PM did, as sometime PM made errors, according to previous experience.

Point 5: deactivate autopilot at final point. Pilot 1 disengaged autopilot at 500 feet (0.15 km) as the runway had been in sight from a far distance and started to fly manually. As Pilot 1 had been away for a while, flying manually from a far distance from the runway was expected to bring back 'the feeling' of controlling the aircraft.

Table 4.6: CDM interview 1: point 5 – deactivate autopilot at final point.

Physical and cognitive demands	Performance	External and internal influences
Pilot 1 turned off autopilot and started to fly the aircraft manually at 500 feet (0.15 km), while kept listening to radio communication with ATC. Pilot 1 mostly looked outside for runway search and PAPI lights, and looked inside to check pitch angle. The aircraft behaviour was also maintained continuously and manually by Pilot 1 so that it follow the glide path.	Runway can be seen far from 500 feet (0.15 km) and auditory announcer gradually told about current altitude. When autopilot was disconnected, auditory feedback can be heard.	Pilot 1 had been a while from flying duty and wanted to 'feel the sensation' again. Procedure from Airbus suggested to land manually under ILS CAT 1; manual textbook also stated about the maximum point for disconnecting autopilot. Combined with experience, these documents shaped Pilot 1 decision-making.

Point 6: touch down. Seconds before touching down on the runway, the aircraft's announcer called out 'retard', indicating the pilot to put the thrust lever to 'idle' position. However, Pilot 1 decided to hold the lever for seconds because according to Pilot 1's judgment, the aircraft will 'fall' due to wind condition. After holding for seconds, Pilot 1 put the lever to 'idle' position and successfully brought the aircraft to runway 25R. While rolling out, Pilot 1 activated the thrust reverser. Pilot 1 mostly looked outside at this stage to maintain the aircraft's position to centre line. The thrust reverser was deactivated by Pilot 1 when the speed reached 70 knots. Before vacating the runway, Pilot 1 switched the radio to 'ground' frequency.

Table 4.7: CDM interview 1: point 6 – touch down.

Physical and cognitive demands	Performance	External and internal influences
When 'retard' called out, Pilot 1 put thrust lever to 'idle' position, after held for seconds for wind correction. After touching down, Pilot 1 activated reverser and mostly looked outside to maintain the aircraft to centre line. When speed reached 70 knots, Pilot 1 deactivated reverser and switched radio to 'ground'.	PM confirmed what had been applied: 'spoilers' to indicate speed brake was active, 'reverse green' to indicate reverser was active, 'decel' to indicate speed was reducing. Pilot 1 could also feel deceleration effect.	'Feeling' and experience shaped decision-making to hold thrust reverser after 'retard' calling.

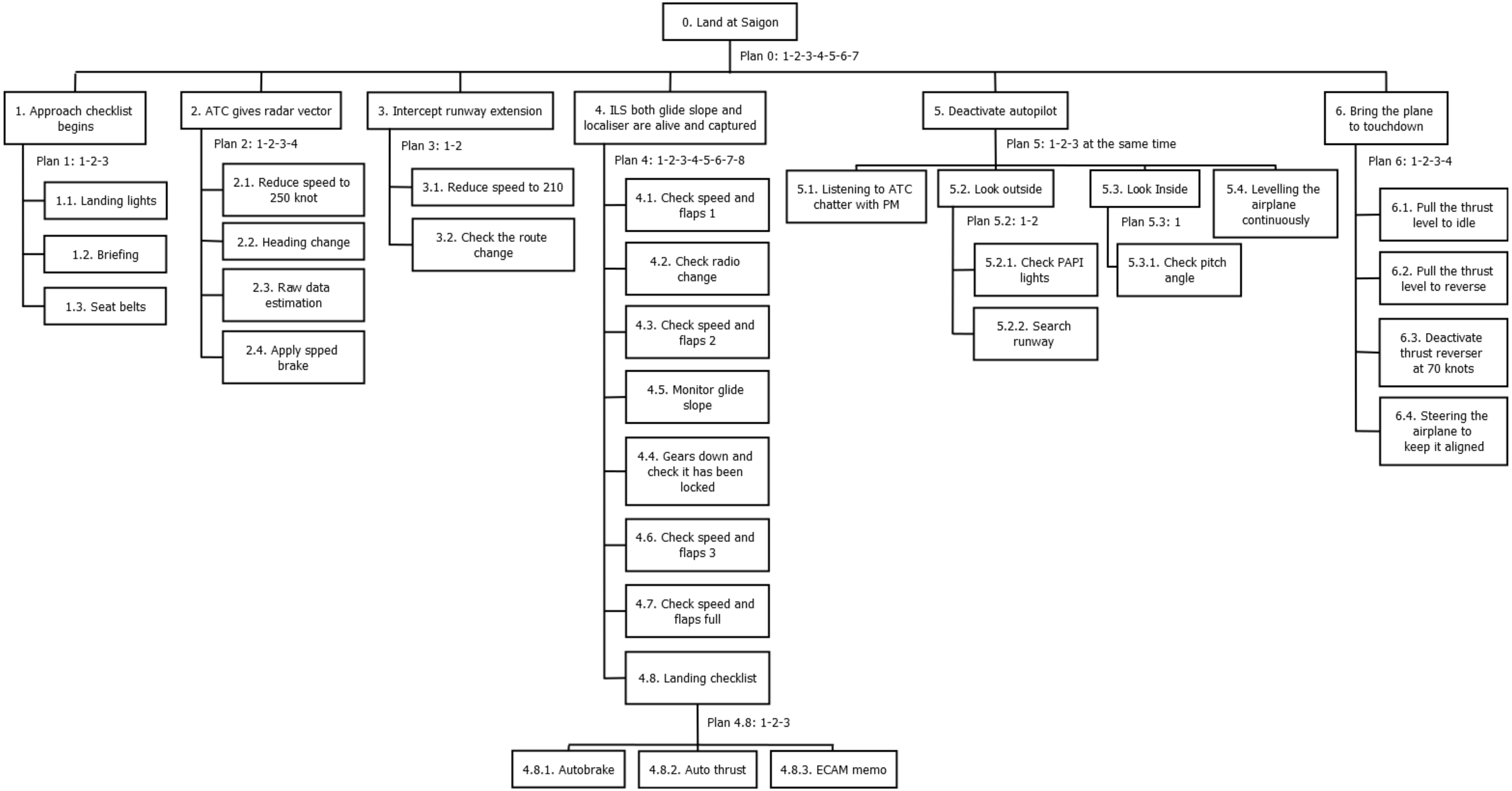


Figure 4.1: Task decomposition of CDM interview 1: landing at Ho Chi Minh City.

4.7.2 CDM Interview 2: Pilot 002

Description

Pilot 2 was a senior Air Force officer with ‘major’ rank. In terms of flying rank, Pilot 2 was a senior captain in the aircraft and was currently flying CN-295. Pilot 2 talked about daily routine training flight around the base. The aircraft took off and landed at the same airfield during daytime. The airfield was actually mixed-used between civilian and military operation. The aircraft was carrying both crew and freight, and Pilot 2 acted as pilot flying (PF).

Pilot 2 considered landing processes was started when the aircraft was leaving holding altitude of 2500 feet (0.76 km) at Alpha 5 point. The aircraft was held at the point for ATC clearance. When cleared, Pilot 2 brought the aircraft towards minimum altitude and set to half-full landing configuration. Pilot 2 sought for the runway while monitoring ATC chatters to get a mental picture about traffic position. After getting clearance to leave holding point, Pilot 2 asked pilot monitoring (PM) to contact ATC.

When reaching a minimum altitude of 400 feet (0.12 km), Pilot 2 sought the runway while checking visibility information and ILS frequency. Pilot 2 maintained the aircraft’s position to the centre of the runway and also continuously checked speed and altitude. While preparing for landing, Pilot 2 also had a plan in mind to go around if something unexpected occurred at this point. When the runway was visually confirmed, Pilot 2 decided to continue landing and set flaps to full position, followed by disconnecting autopilot.

Pilot 2 aligned the aircraft to the runway while struggling with changing winds. Pilot 2 reduced speeds gradually and pulled the control column to make the aircraft ‘flaring out’ seconds before touching down the runway. After all three gears had touched the runway, PF reversed the engines to slow down, and applied braking to get taxi speed. PF vacated the runway after the aircraft reached taxi speed condition and after given instruction from ATC.

A task decomposition diagram of the steps for landing at air force base airfield was developed with Pilot 2, see Figure 4.2 at the end of this section.

Workload points

There are three workload points identified and discussed with Pilot 2, presented in the following tables (Table 4.8, 4.9, and 4.10).

Point 1: leave holding point and start approach. The aircraft was held at Alpha 5 holding area and Pilot 2 was waiting for clearance from ATC. When cleared, the aircraft was descending to minimum altitude point, where Pilot 2 would later decide whether to continue landing or not, while setting the aircraft to half-full landing configuration: flaps approach, landing gears, engines and fuel check, and cabin crew confirmation. At this point, Pilot 2 put focus on outside the aircraft to seek runway. Even though ATC communication was not the main responsibility, Pilot 2 also listened to ATC chatters to draw a mental picture about the aircraft's position and its surrounding traffic. Pilot 2 asked PM to contact tower when leaving holding point Alpha 5.

Table 4.8: CDM interview 2: point 1 – leave holding point and start approach.

Physical and cognitive demands	Performance	External and internal influences
Pilot 2 started approach when leaving holding point Alpha 5. The main tasks at this point were to set approach configuration, seek runway, listen to ATC chatter to draw mental imagery about traffic position, and then ask PM to call ATC when leaving holding point.	PM confirmed PF's orders, and they also appeared in corresponding indicators. Cabin crew also confirmed that passengers were ready.	Pilot 2 used past cases and training to assist in decision-making.

Point 2: reach minimum altitude. The goal of this point was to decide whether to continue landing or not. At this point, Pilot 2 continuously sought for the runway, check visibility, and ILS frequency. Pilot 2 also monitored aircraft's position, speed, and altitude. While preparing for landing, Pilot 2 also had the plan to go around if something unexpected occurred at this point. When the runway could be visually confirmed and Pilot 2 decided to continue landing, flaps were set to full position and autopilot was disconnected.

Table 4.9: CDM interview 2: point 2 – reach minimum altitude.

Physical and cognitive demands	Performance	External and internal influences
Pilot 2 sought for runway, checked visibility information from tower, and made sure that ILS frequency matched with the aircraft. While autopilot was still 'on', Pilot 2 put effort to counter the changing winds so that the aircraft is still aligned to the runway. Altitude and speed were also monitored during approach. When the runway was in sight, flaps were set to full and autopilot was disconnected.	ILS frequency matched the aircraft and the aircraft was flying following the ILS path profile. PM also confirmed that the runway had been in sight.	Pilot 2 used manuals for deciding, such as checklist or QRH. Pilot 2 also compared information obtained before approach.

Point 3: touch down the runway. At this point, Pilot 2 kept the aircraft aligned to the runway as corrections were needed due to changing winds. The speed was gradually reduced and put the aircraft on 'flare out' position seconds before touching down the runway. After three gears had touched the runway, engines were reversed and brakes were applied to slow down the aircraft. When reaching taxi speed, the aircraft vacated the runway after getting instructed by ATC.

Table 4.10: CDM interview 2: point 3 – touch down the runway.

Physical and cognitive demands	Performance	External and internal influences
Pilot 2 gradually reduced speeds accordingly and flared out the aircraft. When all three gears had touched the runway, Pilot 2 activated engine reversers and brakes to slow down the aircraft. Pilot 2 brought the aircraft vacating the runway when reaching taxi speed and getting instruction from ATC.	Instruments told Pilot 2 that the blade had been shifted to 'beta' mode, indicating the reversers were active.	Pilot 2 checked windsock for comparing wind information.

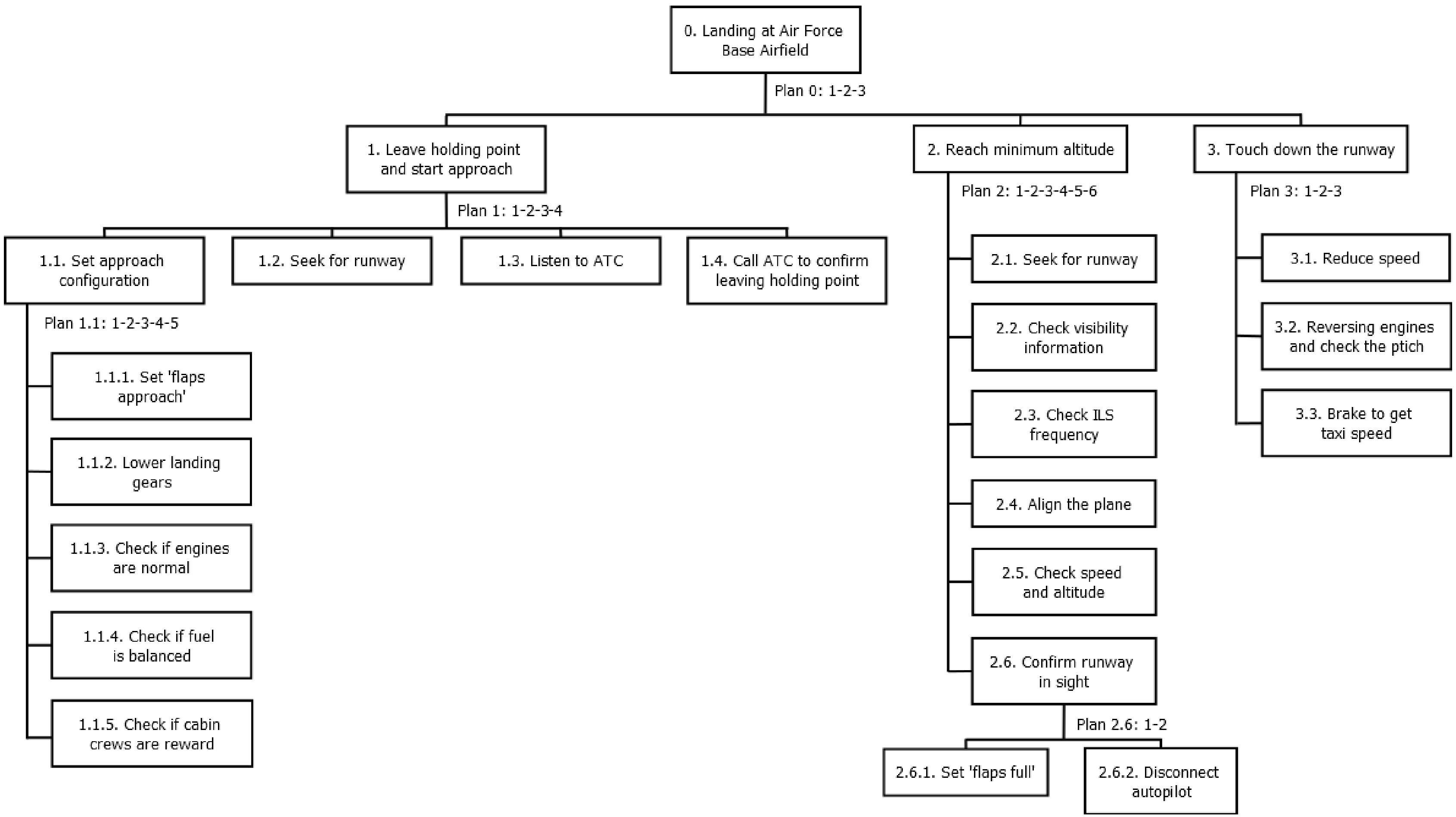


Figure 4.2: Task decomposition of CDM interview 2: landing at air force base airfield.

4.7.3 CDM Interview 3: Pilot 003

Description

Pilot 3 talked about recent flight from Jakarta to Padang City. Pilot 3 flew an Airbus A320 with few passengers due to Covid-19 restrictions. Pilot 3, a senior first officer, acted as pilot flying at this daytime flight. The flight was also an evaluation flight (or check flight) as Pilot 3 had just joined the company recently and completed mandatory flight training.

Pilot 3 considered landing processes started when entering Standard Terminal Arrival (STAR) KATAN 2 Alpha (KATAN2A), 25 nm from the runway, upon clearance from ATC. The airport was located nearby the ocean and the direction to enter the aerodrome was from the south. The aircraft must go slightly to the west while avoiding mountainous areas on the east. The altitude at this point was 8800 feet (2.68 km), the speed was gradually reduced to 250 knots. At 5000 feet (1.52 km), the aircraft passed BAYUR point or initial fixed approach. The distance was 15 nm from the runway. This was a holding point for the airport in case abnormalities occur. The speed continued to decrease to a target speed of 250 knots. Slightly before this point, the approach phase was activated in MCDU.

After BAYUR, there was DME13, 13 nm from the runway. This was the point where the aircraft must be turned to the right to intercept the ILS localiser. The aircraft must be slowed down further. The altitude continued to decrease to achieve target altitude at 3300 feet (1.01 km) later. ILS localiser was expected after that, which were ILS LOC 10.8 nm with the altitude of 3300 feet (1.01 km) and ILS LOC 6 nm with the altitude of 1600 feet (0.49 kilometres). At this point, the aircraft had been cleared for ILS and ready for full landing configuration such as flaps and gears.

Before the runway, there was a Decision Altitude (DA) at 250 feet (76.2 m). At this point, if the runway were not in sight, or something happened on the runway, the aircraft would cancel the landing and go around (missed approach). The autopilot had been turned off as requested by the captain, so Pilot 3 flew the aircraft manually. ATC had cleared the aircraft for landing. As Pilot 3 decided to continue landing, Pilot 3 controlled the aircraft manually by adjusting sink rate until the aircraft touched down the runway. Pilot 3 aimed for touchdown zone markings. Once the aircraft touched the runway, Pilot 3 maintained its position to centre-line and decelerated it using spoilers, brakes, and engine reversers. When the aircraft was about to stop, Pilot 3 transferred control to the captain.

We determined that the steps of landing the aircraft at Padang City were as depicted on a task decomposition diagram (see Figure 4.3 at the end of this section).

Workload points

There are six workload points identified and discussed with Pilot 3, presented in the following tables (Table 4.11, 4.12, 4.13, 4.14, 4.15, and 4.16).

Point 1: pass STAR KATAN2A. Pilot 3 started approach phase when passing STAR KATAN2A. At this point, Pilot 3 did several actions including monitor the flight computer, speed and altitude, autopilot, and traffic both in-front and behind the aircraft. Pilot 3 also reduce speed, set barometric pressure and minimum altitude, initiate approach checklist, and make sure that STAR and runway matched the approach. Pilot 3 also listened to ATC chatters to draw a mental image of surrounding traffic. Flight attendants were told to prepare for landing at this point.

Table 4.11: CDM interview 3: point 1 – pass STAR KATAN2A.

Physical and cognitive demands	Performance	External and internal influences
Upon clearance for ATC for STAR KATAN2A, Pilot 3 initiated several actions including monitoring flight computer, reduce to 250 knots, set local barometric pressure, check altitude, tell flight attendants to prepare landing, check the minimum altitude already set, make sure STAR and runway matched the approach, initiate approach checklist, monitor speed and altitude, autopilot, and surrounding traffic in-front and behind the aircraft. Pilot 3 listened to ATC chatters to make a mental imagery about surrounding traffic.	PM also cross-checked and confirmed for actions to Pilot 3. When the aircraft was behaving as programmed on FMCG, Pilot 3 knew that everything was alright.	Pilot 3 was motivated to know how STAR was conducted in the aerodrome as this was the first time for him flying to this area. The checklist was used as reference for Pilot 3.

Point 2: pass BAYUR point. Passing BAYUR point, Pilot 3 decreased altitude to 5000 feet (1.52 km) and decelerated speed to ‘clean speed’, the lowest speed possible without flaps. Approach phase at MCDU was activated while monitoring inside for speed, altitude, and navigation on instruments. Pilot 3 also monitored outside. At this

point, ‘flap 1’ was set. The aircraft’s flying position and direction were developed in Pilot 3’s mind to assist in decision-making.

Table 4.12: CDM interview 3: point 2 – pass BAYUR point.

Physical and cognitive demands	Performance	External and internal influences
Pilot 3 continued to bring the aircraft to 5000 feet (1.52 km) while reducing speed to ‘clean speed’. Approach phase at MCDU was activated while monitoring aped, altitude, and navigation on instruments. Outside environment was also monitored by Pilot 3. Pilot 3 asked PM to put flap to 1. Pilot 3 developed positions of the aircraft and checked its direction on navigation display.	PF received feedback on monitor confirming that approach phase has been activated. Pilot 3’s display also told about the aircraft’s distance from BAYUR point.	Pilot 3 remembered about SOP. According to Pilot 3’s experience, everything changed a lot despite visiting the same airport every month.

Point 3: pass DME13. Passing DME13 point, Pilot 3 turned the aircraft to intercept ILS localiser, and thus the aircraft must be slowed down. Pilot 3 asked PM for activating ‘flap 2’ while monitoring whether the aircraft was stabilised. Pilot 3 made sure that the target altitude of 3300 feet (1.01 km) was still reachable then turn radio frequency from Padang radar to aerodrome tower frequency.

Table 4.13: CDM interview 3: point 3 – pass DME13.

Physical and cognitive demands	Performance	External and internal influences
The main goal at this point was to prepare for intercepting ILS localiser. So, the aircraft was turned and slowed down while flap was down to 2. Pilot 3 monitored the aircraft behaviour and also thought about the possibility of reaching target altitude. Radio frequency was changed from radar to tower.	n/a	n/a

Point 4: capture ILS LOC 10.8 nm and 6 nm. Around 10.8 nm ILS localiser had been captured and Pilot 3 asked to report ATC. The aircraft continues to decrease its

altitude to 3300 feet (1.01 km). Pilot 3 put the autobrake position to low mode and asked PM to lower the landing gears and made sure that they were up and locked safely by monitoring green lights indicator. Pilot 3 asked for 'flap 3' and immediately 'flap full'. Pilot 3 initiated landing checklists, including checking for landing configuration and display providing feedback of 'landing no blue' status.

Table 4.14: CDM interview 3: point 4 – capture ILS LOC 10.8 nm and 6 nm.

Physical and cognitive demands	Performance	External and internal influences
After cleared by ATC for ILS, the aircraft intercepted ILS localiser while descending to 3300 feet (1.01 km). Pilot 3 put autobrake to low mode and asked PM to lower the landing gears, followed by monitoring lock indicators. Pilot 3 asked for activating 'flap 3' and immediately 'flap full'. Landing checklist was initiated by Pilot 3 and confirmed about landing configuration and 'landing no blue' status.	The display showed 'LOC STAR', and 'GS STAR', then 'GS' if all ILS signals were successfully intercepted. Three green lights and EFIS showed landing gears status if they were down and locked. Subsequently, EFIS displayed a landing configuration page.	n/a

Point 5: reach decision altitude. Before reaching decision altitude, Pilot 3 maintained the aircraft's position to match localiser and glide path, and checked whether configuration for landing has been set. Pilot 3 also checked winds and weather condition. When the runway can be visually confirmed, Pilot 3 decided to continue landing and turned off the autopilot. Pilot 3 monitored speed, altitude, thrust setting, and pitch angle simultaneously and continuously.

Table 4.15: CDM interview 3: point 5 – reach decision altitude.

Physical and cognitive demands	Performance	External and internal influences
Pilot 3 maintained aircraft's position to localiser and glide slope path while checking winds and weather condition. After seeing the runway, Pilot 3 turned off autopilot, while monitored speed, altitude, thrust setting, and pitch angle at the same time.	Pilot 3 could see runway, runway lights, and PAPI lights, indicating the landing plan was correctly executed.	SOP guided Pilot 3 to decide. Pilot 3 had a personal worry about weather phenomena. The captain asked Pilot 3 to do manual landing.

Point 6: touch down. Seconds before touching down, Pilot 3 monitored PAPI lights and reduced sink rate. Pilot 3 landed the aircraft on touch down zone markings, put thrust to idle, and activated engine reversers. Pilot 3 maintained positions of the aircraft to the centre line using rudder pedals. When reaching around 30 knots, thrust reversers were deactivated and Pilot 3 transferred control to the captain.

Table 4.16: CDM interview 3: point 6 – touch down.

Physical and cognitive demands	Performance	External and internal influences
Pilot 3 monitored PAPI lights and reduced sink rate seconds before touching down. Pilot 3 attempted to land the aircraft on touchdown zone markings. When the aircraft touched the runway, Pilot 3 put the thrust lever to idle and then activated engine reversers. Maintaining the aircraft's position to centre line was done by adjusting rudder pedals. When reaching around 30 knots, the thrust lever was brought back to idle and the control of the aircraft was transferred to the captain.	The announcer provided feedback about the aircraft's altitude, saying 'fifty (feet), forty, etc.' Looking outside gave perspective about the aircraft's sink rate. PM confirmed reversers, spoilers, and brakes were working well.	Pilot 3 reminded the training about how to bring the aircraft on zone markings. SOP told Pilot 3 to transferred control to captain because first officers are not allowed to taxi.

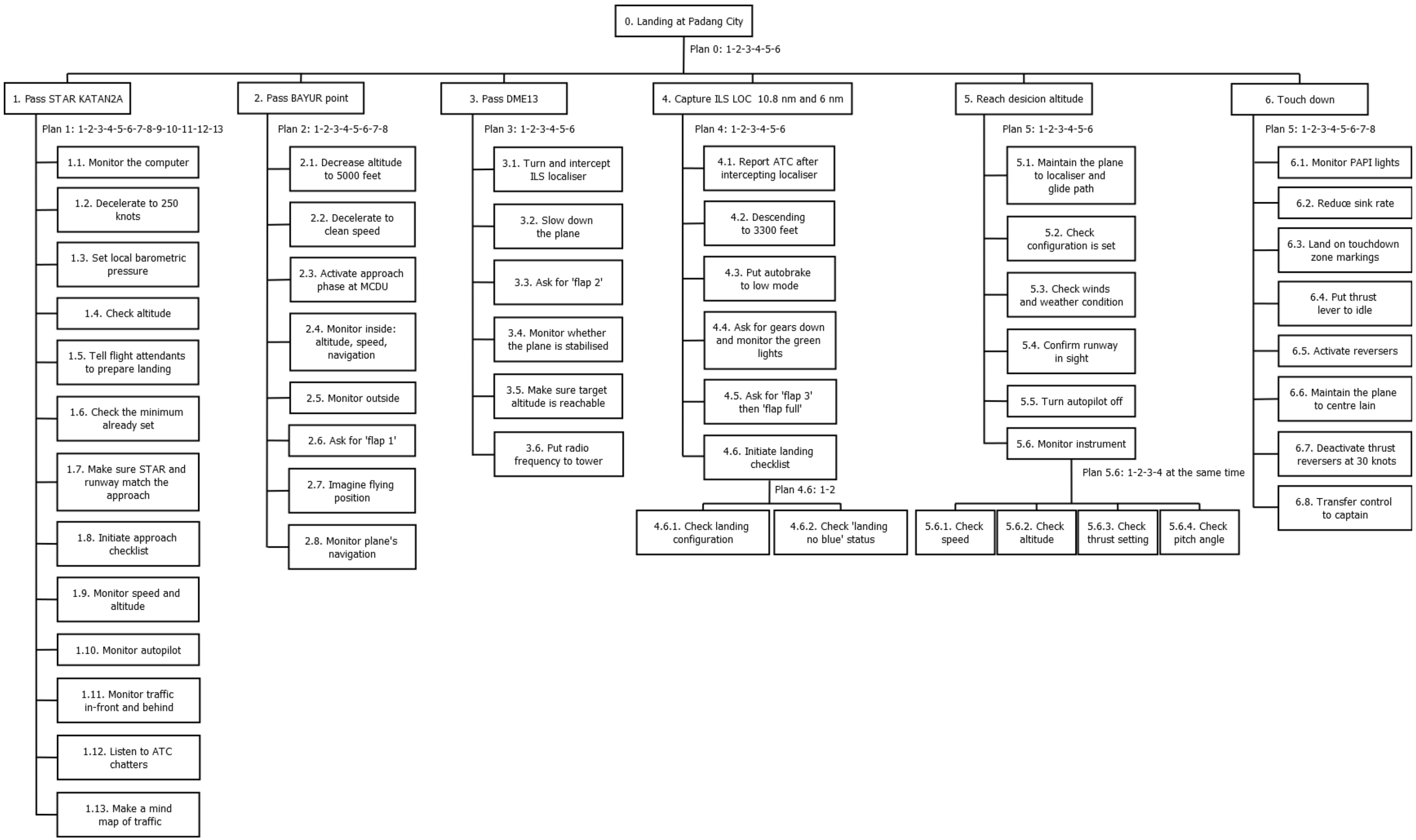


Figure 4.3: Task decomposition of CDM interview 3: landing at Padang City.

4.7.4 CDM Interview 4: Pilot 004

Description

Pilot 4 talked about a recent experience flying an Airbus A330 from Kuala Lumpur in Malaysia to Mumbai in India. The flight was nighttime with almost full passengers. It was in the middle of monsoon season with heavy rain and thunderstorm. Pilot 4, who was a senior flight officer, acted as pilot flying during this flight. The traffic was slightly chaotic that night as most aircraft wanted to avoid the thunderstorm. Pilot 4 had landed in the airport few times.

Pilot 4 considered landing processes were started approximately two hours before actual landing. Pilot 4 asked PM to start landing briefing and calculating numerous parameters for approach. Pilot 4 also listened to ATC communication and weather information to develop a mental image of traffic. After this stage, the aircraft started to descend. Pilot 4 kept an eye for surrounding traffic. Communication with ATC was also intense at this point, as there was an expected delay for traffic entering the airport. The weather was also not good as plenty of thunderstorms occurs in the area.

Pilot 4 started the approach phase according to vector from ATC. Most traffic that night requested for direction changes to avoid thunderstorm. Consequently, there were three changes of approach plan, making Pilot 4 and PM must change and input flight plan on the computer. ATC communication was several times disregarded or slowly responded due to plenty of traffic and weather condition.

Because of fuel concern, Pilot 4 requested immediate vector to join final while keep flying behind slower traffic. At this stage, localiser was captured to guide the aircraft to the centre-line of the runway. Pilot 4 monitored outside for terrain and sought for the runway. The aircraft was maintained at around 4000 feet (1.22 km). After localiser was intercepted, Pilot 4 brought the aircraft to 2900 feet (0.88 km) to capture glide slope. When captured, Pilot 4 check the glide slope and made sure it matched the chart. Final landing configuration such as flaps 'full' and landing gears down were completed at this stage. Pilot 4 had a plan for going around in mind too.

Seconds before touch down, Pilot 4 monitored the sink rate and close the engines power. When the aircraft was rolling out on the runway, Pilot 4 maintained the aircraft's position to centre line using rudders while applying engine reversers. Thereafter, pilot4 monitored brakes and reversers, speed, and also crossing traffic.

Six points has been determined for the scenario of landing at Mumbai, resulting in a task decomposition diagram as seen in Figure 4.4 at the end of this section.

Workload points

There are six workload points identified and discussed with Pilot 4, presented in the following tables: Table 4.17, 4.18, 4.19, 4.20, 4.21, and 4.22.

Point 1: enter 2 hours before landing. At this point, Pilot 4 mostly performed landing briefing and input data for approach plan while listening to ATC and weather information to make sense of surrounding traffic.

Table 4.17: CDM interview 4: point 1 – enter two hours before landing.

Physical and cognitive demands	Performance	External and internal influences
initiating landing briefing was the trigger for this phase. Calculate and plan for approach was also performed, followed by set flight plan to the computer. Listening to ATC and weather information to develop mental imagery of surrounding traffic was also essential at this stage.	n/a	The unwritten rule was used by Pilot 4 to start briefing. Pilot 4 had landed at the airport several times, thus anticipate for cultural issues.

Point 2: start descent. This was the point where the aircraft started to leave its cruising altitude. Pilot 4 monitored descent rate while keeping an eye for traffic and weather. Listening to ATC chatters was also done for developing mental image of surrounding traffic. Pilot 4 asked ATC for expected delay time and execute 'plan A', which was to reduce speed and glide to holding point to save fuel.

Table 4.18: CDM interview 4: point 2 – start descent.

Physical and cognitive demands	Performance	External and internal influences
Monitoring descent rate, traffic, and weather was done while listening to ATC chatters. Pilot 4 had a mental image of traffic based on the information from ATC. Because traffic were plenty at that moment, Pilot 4 asked for expected delay time. Finally, Pilot 4 decided to execute 'plan A': reducing speed and bringing the aircraft to holding point by gliding it.	n/a	The training and experience shaped the way Pilot 4 decided. Pilot 4 also had a thought that humans are inevitably making errors, so Pilot 4 always do cross-check.

Point 3: enter approach phase. This was the point where ATC gave approach plan changes three times, thus Pilot 4 and PM must change flight plan on the computer. These changes happened as consequences of traffic, weather, and, according to Pilot 4 too, culture in managing traffic in the aerodrome. Pilot 4 flew the aircraft by following the airway while always monitoring essential indicators. The aircraft was also descending to 7000 feet (2.13 km).

Table 4.19: CDM interview 4: point 3 – enter approach phase.

Physical and cognitive demands	Performance	External and internal influences
Pilot 4 flew the aircraft to the current airway and complied for restrictions while getting instructions from ATC to change approach plan three times. The aircraft was descending and Pilot 4 monitored fuel, traffic, and weather. While listening to ATC, the aircraft was finally given final approach instruction. Pilot 4 requested delay vector to allow Pilot 4 loading final flight plan.	n/a	Previous experience flying to an area with thunderstorm made Pilot 4 aware that most traffic would request deviation. The understanding of local culture helped Pilot 4 to anticipate instructions. Stories from captain were also beneficial to help decision-making.

Point 4: intercept final. At this stage, the most essential task was to intercept localiser to bring the aircraft centre line to the runway extension. Pilot 4 asked ATC for immediate vector to final due to fuel concern while flying and monitoring slower traffic in-front. Pilot 4 also monitored surrounding terrain, as there was a hill nearby. The aircraft was descending to 4000 feet (1.22 km) when localiser was intercepted. Flaps were extended to 1 and then 2, while seeking for the runway. The speed was reduced to low-speed level of 180 then 170 knots.

Table 4.20: CDM interview 4: point 4 – intercept final.

Physical and cognitive demands	Performance	External and internal influences
Pilot 4 requested immediate vector to final and flew behind the slower traffic in-front. Pilot 4 brought the aircraft to 4000 feet (1.22 km) before intercepting localiser, then applied for flap 1 and immediately flap 2. Pilot 4 kept seeking the runway and reduced speed to 180 then 170 knots. Monitoring outside was also performed as there was a hill nearby.	Indicators on display showed desired feedback. Despite slight chaos, the approach went smoothly.	The chart telling about MSA for the aerodrome helped Pilot 4 to decide. Pilot 4 had knowledge that some airports have restrictions regarding green operation, and this shaped the decision-making too.

Point 5: capture glide slope. The aircraft was descending to 2900 feet (0.88 km) and capture glide slope that bring the aircraft to correct path to the runway. Pilot 4 checked the glide slope against the chart and made sure it matched. The landing gears were lowered, and the flaps were extended to 3 and then full position. Pilot 4 initiated landing checklists and prepare for going around if something unexpected arises.

Table 4.21: CDM interview 4: point 5 – capture glide slope.

Physical and cognitive demands	Performance	External and internal influences
Pilot 4 brought the aircraft to 2900 feet (0.88 km) for capturing glide slope. When active, Pilot 4 checked the chart and made sure the glide slope was correct. Pilot 4 asked for 'landing gears down' and extension of the flap to 3 and immediately to full position. Pilot 4 checked landing configuration by initiating landing checklist, while also preparing for going around in case unexpected events occur.	Indicators told Pilot 4 that everything was OK, and landing configuration was correct. The aircraft performance also indicated that it's ready to land. The runway had been in sight from around 2900 feet (0.88 km).	The knowledge of certain airport regulation helped Pilot 4 to decide.

Point 6: touch down. At this final stage, Pilot 4 monitored sink rate and closed the engine power seconds before touching down. Immediately after rolling out the runway, the next job was to maintain the aircraft's position to centre line using rudder pedals and applied thrust reverser to decelerate. Pilot 4 monitored braking systems and reversers and made sure the speed was decreasing. Because the runways were coring each other, Pilot 4 also concerned about crossing traffic.

Table 4.22: CDM interview 4: point 6 – touch down.

Physical and cognitive demands	Performance	External and internal influences
During final moment, Pilot 4 monitored sink rate thus when touching down the runway the aircraft was in correct position. Engine power was also closed and immediately after rolling out, Pilot 4 maintained position to the centre line of the runway using rudder pedals. Pilot 4 activated and monitored thrust reversers along with braking systems and made sure the speed was decreasing. Pilot 4 also monitored crossing traffic.	n/a	Charts and procedures guided Pilot 4 to implement strategies for the landing.

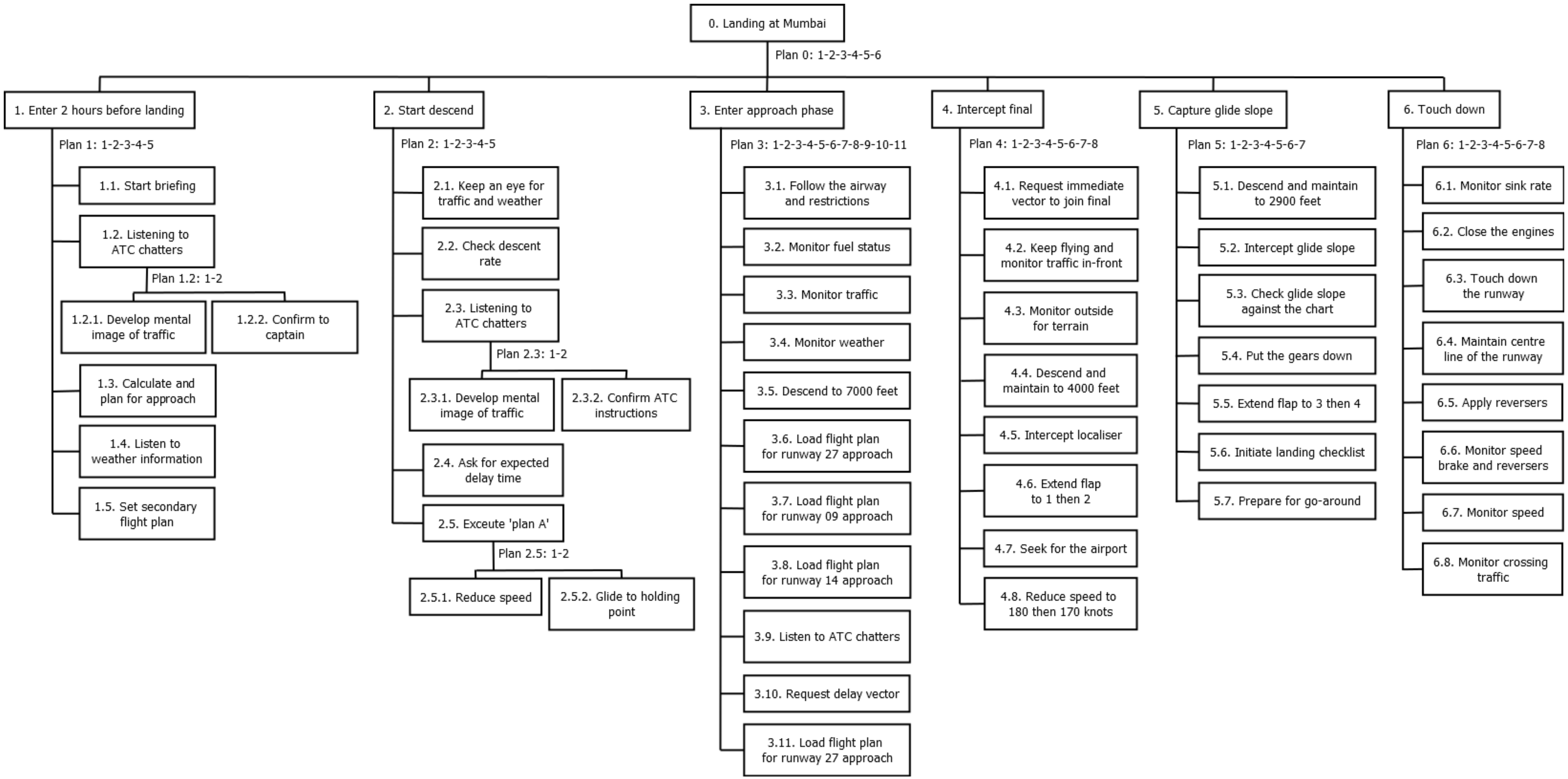


Figure 4.4: Task decomposition of CDM interview 4: landing at Mumbai.

4.7.5 CDM Interview 5: Pilot 005

Description

Pilot 5, a former air force pilot, talked about a recent flight to Manado City (MDC). The aircraft was an Airbus A330-200 with full passengers. It was a daytime flight and Pilot 5 acted as pilot flying. The characteristics of the destination airport were slightly challenging, particularly for the wide body airliner: surrounded by mountains, short and narrow runways, and limited navigational aids. Pilot 5's rank in the current company was a senior flight officer.

Pilot 5 considered landing processes were started at ToD or 'Top or Descend', around 100 nm from the airport. At this stage, PF asked PM to seek information or data about the destination aerodrome. Pilot 5 used the old data obtained from departure while waiting for the updated one. After passing LUANG point, which was an initial fix for STAR, Pilot 5 did not follow STAR and requested for heading change to avoid thunderstorm. Thereafter, the aircraft entered terminal aerodrome at 25 nm before the airport. ATC asked Pilot 5 to hold the aircraft at 10,000 feet (3.05 km) while preparing landing configurations such as autobrake, flaps, and maintaining normal speed for this phase.

At initial approach fix, the aircraft was held and circling the holding pattern. Pilot 5 performed a small readjustment as the aircraft was flying in the middle of a mountainous area. Entering intermediate approach fix, the aircraft had been fully configured for landing. Pilot 5 slightly circled the airport before getting to the approach fix. Finally, Pilot 5 entered final approach fix and saw the runway lights at around 800 feet (0.24 km), thus decided to continue landing. Glide slope and localiser had been captured and full flaps configuration had been made. The aircraft touched the runway and decelerated uneventfully.

Six points has been determined for the scenario of landing at Manado City, resulting in a task decomposition diagram as seen in Figure 4.5 at the end of the section.

Workload points

There are six workload points identified and discussed with Pilot 5, presented in the following tables: Table 4.23, 4.24, 4.25, 4.26, 4.27, and 4.28.

Point 1: enter ToD. At this point, Pilot 5 sought information about weather (ATIS) and aerodrome. Landing briefing was initiated while inputting old data from departure airport and set autobrake. After getting instruction from ATC, new data was inputted. Pilot 5 confirmed and checked the data and adjust descent profile of the aircraft.

Table 4.23: CDM interview 5: point 1 – enter ToD.

Physical and cognitive demands	Performance	External and internal influences
Pilot 5 asked PM to seek information on aerodrome and weather around. Old data from departure airport was used as initial reference and set autobrake. ATC instructed preparing landing at runway 36, so Pilot 5 changed the old data with the updated one. Pilot 5 confirmed and checked the data before executing it. The aircraft descend profile was quickly adjusted by Pilot 5 to match the new plan.	100 nm ToD had been included in FMCG.	Experience and training helped Pilot 5 to act during landing preparation, assisted by NOTAM, manuals, and also a calculation formula shaped by experience.

Point 2: pass LUANG point. At this point, Pilot 5 decided to fly without following standard route to avoid the thunderstorm while continuing descent. Pilot 5 updated the flight plan and check the data against flight performance. Pilot 5 also turned the aircraft to fly north-east to get an upwind position. Pilot 5 always updated aircraft's position by confirming radial and distance to ATC, as MDC ATC did not have radar contact. Mental image of traffic was also developed by Pilot 5.

Table 4.24: CDM interview 5: point 2 – pass LUANG point.

Physical and cognitive demands	Performance	External and internal influences
Pilot 5 tried to find an escape route from thunderstorm ahead by contacting ATC to reject STAR and request for heading. The aircraft was descending, and Pilot 5 updated the flight plan and checked the data against flight performance. Pilot 5 realised the wind condition and turned the aircraft to fly north-east to get an upwind position. As Pilot 5 flew manually and ATC had no radar contact, Pilot 5 updated position by confirming radial and distance. Mental image of traffic was also developed by Pilot 5 at this point.	n/a	n/a

Point 3: enter terminal aerodrome. This point was crucial as Pilot 5 was flying in the middle of mountains. Pilot 5 continued to descend to 8600 feet (2.62 km) as recommended by the chart for MSA of the area, and then passed the north radial to continue descent to 7000 feet (2.13 km). Pilot 5 set autobrake and maintained speed to 230-250 knots. Pilot 5 started to configure flaps for landing.

Table 4.25: CDM interview 5: point 3 – enter terminal aerodrome.

Physical and cognitive demands	Performance	External and internal influences
Pilot 5 continued descent carefully following MSA for this area as written on the chart, then passed the north radial and descended to 7000 feet (2.13 km). Autobrake had been set and Pilot 5 maintained the speed around 230-250 knots. Pilot 5 then started to configure flaps for landing.	n/a	Company policy helped Pilot 5 to decide when to lower the flaps.

Point 4: initial approach fix. At this point, Pilot 5 was asked to hold the aircraft at the holding pattern, which was in the middle of the mountain. Pilot 5 selected flap to position 1 and applied low-speed flying. Pilot 5 kept listening to ATC and confirmed when another traffic had landed safely.

Table 4.26: CDM interview 5: point 4 – initial approach fix.

Physical and cognitive demands	Performance	External and internal influences
Pilot 5 flew the aircraft to holding pattern as instructed by ATC. Pilot 5 selected flap 1 and flew in low-speed mode. Pilot 5 listened to ATC chatter and confirmed when another traffic had landed safely.	n/a	n/a

Point 5: intermediate approach fix. Pilot 5 started to configure the aircraft for full landing configuration. Pilot 5 also monitor surrounding terrain and the speed as the aircraft was flying at low-speed mode. Pilot 5 selected flap 2 and immediately flap 3, and then lowered the landing gears.

Table 4.27: CDM interview 5: point 5 – intermediate approach fix.

Physical and cognitive demands	Performance	External and internal influences
Pilot 5 configured the aircraft for full landing while monitored for terrain and low-speed flying. Pilot 5 selected flap to 2 and immediately 3, and then lowered the landing gears.	n/a	n/a

Point 6: final approach fix. At this final moment, Pilot 5 had prepared a plan for going around in case of emergency. Pilot 5 saw the runway lights at around 800 feet (0.24 km), thus decided to continue landing and disconnected the autopilot. For a couple of seconds, Pilot 5 tried to feel the control of the aircraft while focusing on the outside environment. Pilot 5 focused to listen to automatic call-outs and cut the power when called ‘retard’. Pilot 5 put the aircraft on the runway and maintained centre line using rudder pedals. Pilot 5 applied and monitored deceleration methods to bring the aircraft vacating the runway.

Table 4.28: CDM interview 5: point 6 – final approach fix.

Physical and cognitive demands	Performance	External and internal influences
While focusing on landing the aircraft, Pilot 5 had to think about the plan to go around in case of emergency. Pilot 5 confirmed runway in sight at 800 feet (0.24 km) and disconnected autopilot at 500 feet (0.15 km). Pilot 5 tried to feel the control of the aircraft for a while and focused on the outside environment. Pilot 5 listened to automatic call-outs from the aircraft and cut the power when called ‘retard’. Pilot 5 brought the aircraft to the runway and maintained to centre line using rudder pedals. Pilot 5 also applied and monitored deceleration methods to slow down the aircraft.	When disconnecting autopilot, the aircraft position was automatically at its last state. Seeing the runway light was the good feedback indicating the aircraft was on the right track. PM told Pilot 5 that deceleration methods worked well.	Manual from Airbus helped Pilot 5 to decide when to use autopilot.

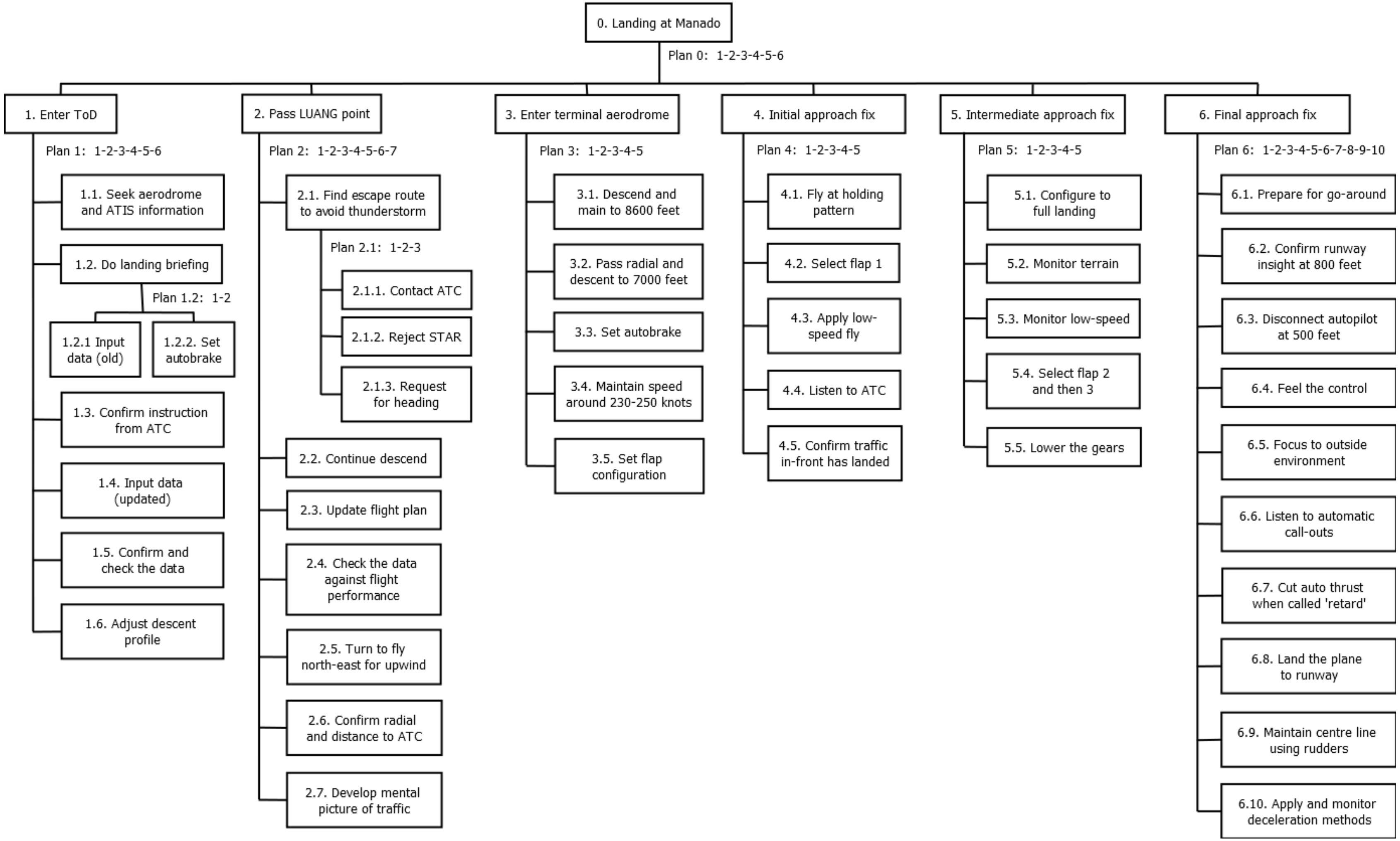


Figure 4.5: Task decomposition of CDM interview 5: landing at Manado City

4.8 Summary from the CDM Interviews

After classifying tasks into the mental workload framework from five CDM interviews, we concluded there are several common actions that pilots must complete when landing an aircraft. These actions, listed below, may act as possible sources of physical and cognitive demands that later create mental workload. Table 4.29 compiles evidence from CDM interviews.

1. **Prepare data for approach**: Here pilots sought information for the destination airport such as runway, weather, and so forth. The action at this point included inputting the data to the flight computer and performing calculations.
2. **Communicate with ATC**: Pilots always communicated with ATC for updates about the destination airports such as active runway, wind speed and direction, and so forth. Here, pilots also listened to ATC chatters in the frequency to develop a mental image about their aircraft's position and surrounding traffic.
3. **Update flight plan if any**: Often pilots must update plans for landing because of changes in the way pilots must approach the aerodrome or changes in active runways. ATC frequently gave vectors or directions to intercept final points, and thus pilots must update the flight plan and computer as instructed.
4. **Monitor outside environment**: When the aircraft was descending and flying at lower altitude, pilots must also look outside the aircraft to seek for the runway and traffic. This was even more important if potential obstacles such as hills or mountains exists nearby the destination airport.
5. **Configure the aircraft for landing**: Along the way when the approach phase had been started, pilots must configure the plane so that it is ready for landing. The configuration must be made according to the certain criteria such as altitude, speed, and also company policy. The actions were including setting autobrake, selecting flaps, and lowering landing gears.
6. **Fly and land the aircraft manually**: It was common for pilots, and sometimes recommended by company or manufacturer, to land the aircraft manually after confirming the runway was in sight. Currently, pilots thoroughly controlled the aircraft movement laterally and vertically, assisted by flight computer system for speed adjustment and navigation (ILS). Upon touching down and rolling out, pilots attempted to decelerate by applying and monitoring speed brake, spoilers, and reversers while maintaining the aircraft's position to runway centre line using rudder pedals.

Table 4.29: Evidence for possible sources of physical and cognitive demands.

Sources	Evidence
Prepare data for approach	Altimeter 10,000 feet (3.05 km) was used as a clue for initiating approach checklist (Pilot 001). PM read the checklist items while Pilot 1 responded with actions and confirmed. Pilot 1 also performed rough estimation and calculation, concluded that the aircraft would be higher when intercepting, and thus applied speed brake (Pilot 001). Pilot 2 sought for runway, checked visibility information from tower, and made sure that ILS frequency matched with the aircraft (Pilot 002). Upon clearance for ATC for STAR KATAN2A, Pilot 3 initiated several actions including monitoring flight computer, reduce to 250 knots, set local barometric pressure, check altitude, tell flight attendants to prepare landing, check the minimum altitude already set, make sure STAR and runway matched the approach, initiate approach checklist, monitor speed and altitude, autopilot, and surrounding traffic in-front and behind the aircraft (Pilot 003). Initiating landing briefing was the trigger for this phase. Calculate and plan for approach was also performed, followed by set flight plan to computer (Pilot 004). Pilot 5 asked PM to seek information on aerodrome and weather around. Old data from departure airport was used as initial reference and set autobrake (Pilot 005).
Communicate with ATC	ATC asked to descend to 3000 feet (0.91 km) from FL410 (Pilot 001). Pilot 1 listened to radio communication between PM and ATC, and made heading changes as instruction and confirmed (Pilot 001). Pilot 1 turned off autopilot and started to fly the aircraft manually at 500 feet (0.15 km), while kept listening to radio communication with ATC (Pilot 002). Pilot 3 listened to ATC chatters to make a mental imagery about surrounding traffic (Pilot 003). Listening to ATC and weather information to develop mental imagery of surrounding traffic was also essential at this stage (Pilot 004). Pilot 4 had a mental image of traffic based on the information from ATC (Pilot 004). While listening to ATC, the aircraft was finally given final approach instruction (Pilot 004). Pilot 4 requested delay vector to allow Pilot 4 loading final flight plan (Pilot 004). Pilot 4 requested immediate vector to final and flew behind the slower traffic in-front (Pilot 004). ATC instructed preparing landing at runway 36, so Pilot 5 changed the old data with the updated one (Pilot 005). Pilot 5 tried to find an escape route from thunderstorm ahead by contacting ATC to reject STAR and request for heading (Pilot 005). Pilot 5 updated position by confirming radial and distance (Pilot 005). Pilot 5 listened to ATC chatter and confirmed when another traffic had landed safely (Pilot 005).

(continued on next page)

Table 4.29, continued

Sources	Evidence
Update flight plan if any	Pilot 1 received confirmation of route change from PM and checked it on FMCG (Pilot 001). Because traffic were plenty at that moment, Pilot 4 asked for expected delay time (Pilot 004). Finally, Pilot 4 decided to execute 'plan A': reducing speed and bringing the aircraft to holding point by gliding it (Pilot 004). Pilot 4 flew the aircraft to the current airway and comply for restrictions while getting instructions from ATC to change approach plan three times (Pilot 004). While listening to ATC, the aircraft was finally given final approach instruction. Pilot 4 requested delay vector to allow Pilot 4 loading final flight plan (Pilot 004). ATC instructed preparing landing at runway 36, so Pilot 5 changed the old data with the updated one (Pilot 005). The aircraft descend profile was quickly adjusted by Pilot 5 to match the new plan (Pilot 005).
Monitor outside environment	Pilot 1 mostly looked outside for runway search and PAPI lights, and looked inside to check pitch angle (Pilot 001). Pilot 2 sought for runway, checked visibility information from tower, and made sure that ILS frequency matched with the aircraft (Pilot 002). Outside environment was also monitored by Pilot 3 (Pilot 003). Pilot 4 kept seeking the runway and reduced speed to 180 then 170 knots (Pilot 004). Monitoring outside was also performed as there was a hill nearby (Pilot 004). Pilot 5 continued descent carefully following MSA for this area as written on the chart, then passed the north radial and descended to 7000 feet (2.13 km) (Pilot 005). Pilot 5 configured the aircraft for full landing while monitored for terrain and low-speed flying (Pilot 005). Pilot 5 tried to feel the control of the aircraft for a while and focused on the outside environment (Pilot 005).
Configure the aircraft for landing	Pilot 1…applied speed brake (Pilot 001). Pilot 1 asked PM to lower flaps from 'up' position to 'full' one-by-one according to the speed (Pilot 001). When the runway was in sight, flaps were set to full and autopilot was disconnected (Pilot 002). Pilot 3 asked PM to put flap to 1 (Pilot 003). So, the aircraft was turned and slowed down while flap was down to 2 (Pilot 003). Pilot 3 put autobrake to low mode and asked PM to lower the landing gears, followed by monitoring lock indicators (Pilot 003). Pilot 3 asked for activating 'flap 3' and immediately 'flap full' (Pilot 003). Pilot 4 brought the aircraft to 4000 feet (1.22 km) before intercepting localiser, then applied for flap 1 and immediately flap 2 (Pilot 004). Pilot 4 asked for 'landing gears down' and extension of the flap to 3 and immediately to full position (Pilot 004). Autobrake had been set and Pilot 5 maintained the speed around 230-250 knots (Pilot 005). Pilot 5 then started to configure flaps for landing (Pilot 005). Pilot 5 selected flap 1 and flew in low-speed mode (Pilot 005). Pilot 5 selected flap to 2 and immediately 3, and then lowered the landing gears (Pilot 005).

(continued on next page)

Table 4.29, continued

Sources	Evidence
Fly and land the aircraft manually	Pilot 1 turned off autopilot and started to fly the aircraft manually at 500 feet (0.15 km), while kept listening to radio communication with ATC (Pilot 001). The aircraft behaviour was also maintained continuously and manually by Pilot 1 so that it follow the glide path (Pilot 001). When ‘retard’ called out, Pilot 1 put thrust lever to ‘idle’ position, after held for seconds for wind correction (Pilot 001). After touching down, Pilot 1 activated reverser and mostly looked outside to maintain the aircraft to centre line (Pilot 001). When speed reached 70 knots, Pilot 1 deactivated reverser and switched radio to ‘ground’ (Pilot 001). When the runway was in sight, flaps were set to full and autopilot was disconnected (Pilot 002). Pilot 2 gradually reduced speeds accordingly and flared out the aircraft (Pilot 002). When all three gears had touched the runway, Pilot 2 activated engine reversers and brakes to slow down the aircraft (Pilot 002). After seeing the runway, Pilot 3 turned off autopilot, while monitored speed, altitude, thrust setting, and pitch angle at the same time (Pilot 003). Pilot 3 monitored PAPI lights and reduced sink rate seconds before touching down (Pilot 003). Pilot 3 attempted to land the aircraft on touchdown zone markings (Pilot 003). When the aircraft touched the runway, Pilot 3 put the thrust lever to idle and then activated engine reversers (Pilot 003). Maintaining the aircraft’s position to centre line was done by adjusting rudder pedals (Pilot 003). During final moment, Pilot 4 monitored sink rate, thus when touching down the runway the aircraft was in correct position (Pilot 004). Engine power was also closed and immediately after rolling out, Pilot 4 maintained position to the centre line of the runway using rudder pedals (Pilot 004). Pilot 4 activated and monitored thrust reversers along with braking systems and made sure the speed was decreasing (Pilot 004). Pilot 5 confirmed runway in sight at 800 feet (0.24 km) and disconnected autopilot at 500 feet (0.15 km) (Pilot 005). Pilot 5 tried to feel the control of the aircraft for a while and focused on the outside environment (Pilot 005). Pilot 5 listened to automatic call-outs from the aircraft and cut the power when called ‘retard’ (Pilot 005). Pilot 5 brought the aircraft to the runway and maintained to centre line using rudder pedals (Pilot 005). Pilot 5 also applied and monitored deceleration methods to slow down the aircraft (Pilot 005).

Every action that pilots do will result in performance feedback. From CDM interviews, we find three distinct types of feedback, as listed below. The evidence of CDM interviews for this part are shown in Table 4.30.

1. **Feedback is shown on display or auditory call-out**: Pilots were getting feedback from aircraft system through display or visual and auditory devices. From this kind of feedback, pilots would realise whether their actions result in desired effect and might influence workload.

2. **Feedback is concluded from physical changes in the aircraft and environment**: Pilots also could see or feel the effect of their actions through changes in physical movement of the aircraft or outside environment, such as feeling of deceleration, seeing the runway, etc. This feedback also indicated that the actions resulted in desired effect and may influence workload.

3. **Feedback is confirmed by pilot monitoring**: Here, pilots were told by their flying partner about the actions they had just done. Conformation from pilot monitoring (PM) served as feedback that their actions were correct and resulted in desired effect. This kind of feedback might also affect the way pilots perceive workload.

Table 4.30: Evidence for performance feedback.

Feedback	Evidence
Feedback are shown on display or auditory call-out	The aircraft was slightly too high as predicted, and the route change was displayed on monitor (Pilot 001). All technical actions feedback were shown on the display (Pilot 001). When autopilot was disconnected, auditory feedback can be heard (Pilot 001). Runway can be seen far from 500 feet (0.15 km) and auditory announcer gradually told about current altitude (Pilot 001). ILS frequency matched the aircraft and the aircraft was flying following the ILS path profile (Pilot 002). Instruments told Pilot 2 that the blade had been shifted to ‘beta’ mode, indicating the reversers were active (Pilot 002). PF received feedback on monitor confirming that approach phase has been activated (Pilot 003). Pilot 3’s display also told about the aircraft’s distance from BAYUR point (Pilot 003). The display showed ‘LOC STAR’, and ‘GS STAR’, then ‘GS’ if all ILS signals were successfully intercepted (Pilot 003). Three green lights and EFIS showed landing gears status if they were down and locked (Pilot 003). Thereafter, EFIS displayed landing configuration page (Pilot 003). The announcer provided feedback about the aircraft’s altitude, saying ‘fifty (feet), forty, etc.’ (Pilot 003). Indicators on display showed desired feedback (Pilot 004). Indicators told Pilot 4 that everything was OK, and landing configuration was correct (Pilot 004). 100 nm ToD had been included in FMCG (Pilot 005).

(continued on next page)

Table 4.30, continued

Feedback	Evidence
Feedback are concluded from physical changes in the aircraft and environment	Runway can be seen far from 500 feet (0.15 km) and auditory announcer gradually told about current altitude (Pilot 001). Pilot 1 also could feel deceleration effect (Pilot 001). When the aircraft was behaving as programmed on FMCG, Pilot 3 knew that everything was alright (Pilot 003). Looking outside gave perspective about the aircraft's sink rate (Pilot 003). The aircraft performance also indicated that it's ready to land (Pilot 004). Runway had been in sight from around 2900 feet (0.88 km) (Pilot 004). Seeing the runway light was the good feedback indicating the aircraft was on the right track (Pilot 005).
Feedback is confirmed by pilot monitoring	PM confirmed what had been applied: 'spoilers' to indicate speed brake was active, 'reverse green' to indicate reverser was active, 'decel' to indicate speed was reducing (Pilot 001). PM confirmed PF's orders, and they also appeared in corresponding indicators (Pilot 002). Cabin crew also confirmed that passengers were ready (Pilot 002). PM also confirmed that runway had been in sight (Pilot 002). PM also cross-checked and confirmed for actions to Pilot 3 (Pilot 003). PM confirmed reversers, spoilers, and brakes were working well (Pilot 003). PM told Pilot 5 that deceleration methods worked well (Pilot 005).

External and internal influences may affect the way pilots perceive workload. From CDM interviews, included in this aspect were motivation, experience, training, company policies, manufacturer recommendations, and cultures. The evidence of CDM interviews for this part are shown in Table 4.31.

1. **Influence from manuals or policies**: Here written or unwritten rules may shape the way pilots decide to perform certain actions during landing thus might affect the way they see their workload.

2. **Influence from experience or training**: Previous experience flying to the same airport or having been trained for certain scenarios may affect workload.

Table 4.31: Evidence for external and internal influences.

Influences	**Evidence**
Influence from manuals or policies	Reducing from 250 to 210 was a normal procedure (Pilot 001). Pilot 1 also had MSA as reference for surrounding terrain (Pilot 001). Procedure from Airbus suggested to land manually under ILS CAT 1; manual textbook also stated about the maximum point for disconnecting autopilot (Pilot 001). Pilot 2 used manuals for deciding such as checklist or QRH (Pilot 002). Pilot 2 also compared information obtained before approach (Pilot 002). Checklist was used as reference for Pilot 3 (Pilot 003). Pilot 3 remembered about SOP (Pilot 003). The unwritten rule was used by Pilot 4 to start briefing (Pilot 004). MSA on the chart for the aerodrome helped Pilot 4 to decide (Pilot 004). Charts and procedures guided Pilot 4 to implement strategies for the landing (Pilot 004). Experience and training helped Pilot 5 to act during landing preparation, assisted by NOTAM, manuals, and also a calculation formula shaped by experience (Pilot 005). Company policy helped Pilot 5 deciding when to lower flaps (Pilot 005). Airbus manual helped Pilot 5 deciding when to use autopilot (Pilot 005).
Influence from experience or training	Pilot 1 had memorised the checklist item as it has been done plenty of times (Pilot 001). According to Pilot 1's experience on past flights, if ATC directs the aircraft directly to near the runway, the altitude will be higher (Pilot 001). Pilot 1 checked what PM did, as sometime PM made errors, according to previous experience (Pilot 001). Pilot 1 had been a while since the last flight and wanted to 'feel the sensation' again (Pilot 001). 'Feeling' and experience shaped decision-making to hold thrust reverser after 'retard' calling (Pilot 001). Pilot 2 used past cases and training to assist in decision-making (Pilot 002). Pilot 2 checked windsock for comparing wind information (Pilot 002). Pilot 3 was motivated to know how STAR was conducted in the aerodrome as this was the first time for him flying to this area (Pilot 003). According to Pilot 3's experience, everything changed a lot despite visiting the same airport every month (Pilot 003). Pilot 4 had landed at the airport several times, thus anticipate for cultural issues (Pilot 004). Training and experience shape the way Pilot 4 decided (Pilot 004). Pilot 4 also had a thought that humans are inevitably making errors, so Pilot 4 always do cross-check (Pilot 004). Previous experience flying to an area with thunderstorm made Pilot 4 aware that most traffic would request deviation (Pilot 004). The understanding of local culture helped Pilot 4 to anticipate instructions (Pilot 004). Stories from the captain were also beneficial to help decision-making (Pilot 004). Pilot 4 had knowledge that some airports have restrictions regarding green operation, and this shaped the decision-making too (Pilot 004). The knowledge of certain airport regulation helped Pilot 4 to decide (Pilot 004). Experience and training helped Pilot 5 to act during landing preparation, assisted by NOTAM, manuals, and also a calculation formula shaped by experience (Pilot 005).

4.9 Discussion

This study aims to explore pilots' mental workload during landing processes using a qualitative approach. It has been seemingly widely accepted that landing is one of the most demanding parts of a flight. However, what generates demands and the way they interact with performance feedback and external/internal influences during this crucial stage may need to be explored further. CDM interviews serve as a tool to elicit expert knowledge about the way they complete certain tasks based on their real experience. Even though pilots might have faced similar problems during their careers, the way they approach the problems tend to be peculiar to each individual flight.

From CDM interviews, we identified six distinct actions that commonly appear during the landing stage that could be the source of mental and physical demands. We also identified three sources of performance feedback and two sources of external or internal influences that interact on each other to form mental workload. Figure 4.6 shows the relationship diagram portraying sources of demands, feedback, and influences inferred from CDM interviews.

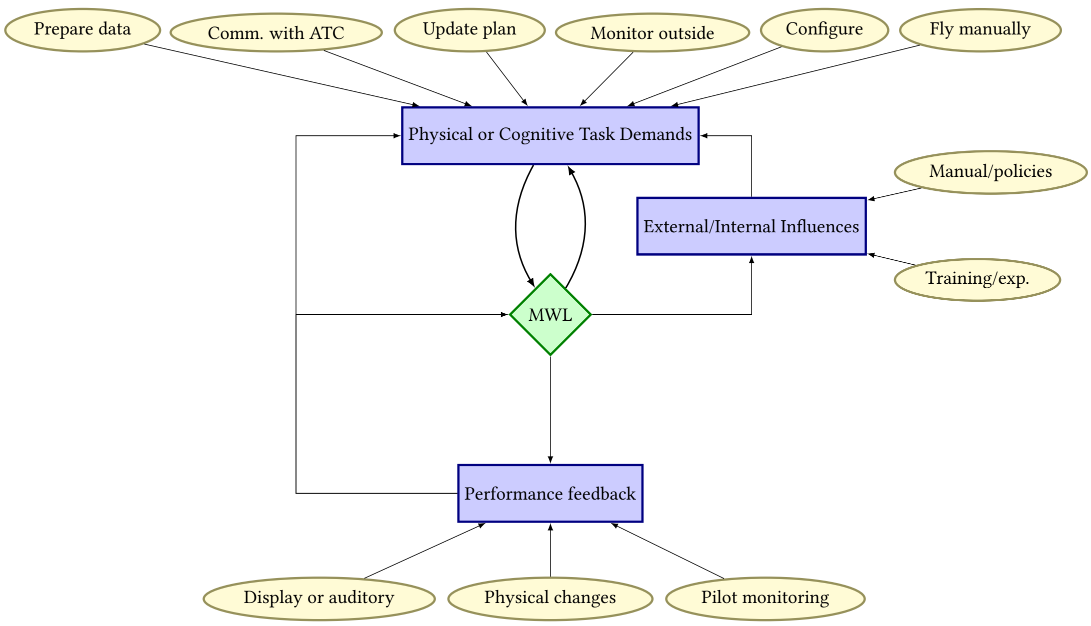


Figure 4.6: Sources of task demands, performance feedback, and influences within MWL framework based on CDM interviews.

Data preparation was, in general, performed when pilots decide to initiate the approach phase, which is the starting point of the landing processes. There were several clues to start this stage: altitude (Pilot 1) or certain navigational points (Pilot 2, 3, and 4), or time (Pilot 5), depending on company or manufacturer manuals. At this point, pilots attempted to obtain data of destination airport and set them to the flight computer. They usually had obtained the data from the departure airport, however, situations on the destination airport might dynamically change. Pilots may also use their calculation to produce 'raw data' as initial reference, while waiting for updated one from ATC.

The next action was communicating with ATC. It had been obvious that pilots always maintain communication with ATC. It was actually the main duty of pilot monitoring yet pilot flying also took responsibility to listen to the radio. The way pilots did communications with ATC can be classified into two forms. First, pilots communicated directly for their flight. This includes reporting their position, complying to ATC instruction, or requesting and reporting for certain manoeuvrings. Secondly, pilots paid attention to other traffic's communication with ATC. Although ATC, for instance, gave instruction to another aircraft, pilots usually noticed this communication as well. This was considered essential to maintain their situational awareness by drawing a mental picture of traffic position. They had to know the position of their aircraft and other traffic to anticipate if something unusual occurred.

Updating a plan was common during the approach phase and can be possibly the source of high mental demands. Pilots expected to get updated data of their destination airport from ATC. When an updated one had been received, it was the pilots' job to input the data to the flight computer. However, this could occur several times due to several factors. Pilot 4, for instance, discussed his experience of getting the approach changed four times. At that time of flying, the aircraft had a concern regarding its fuel quantity, thus being vectored by ATC to match a changing approach plan was a demanding experience. Pilot 5 mentioned similar experience when he had to change the plan twice because active runway was changed due to wind conditions.

When reaching low altitude during an approach phase, a pilot might see the outside environment and start to build awareness of its surrounding. For example, pilots might have known from the chart that the destination airport was surrounded by hills (Pilot 5). However, pilots needed assurance that they were flying on the correct path, and they were not flying towards the hills. One of the ways to build this assurance was by confirming visually. Monitoring the outside environment was also helpful to land the aircraft. Precision Approach Path Indicator (PAPI) lights, that are located on the left of the runway, can guide the pilot to the correct path of landing. Visually confirming the runway, or the runway lights, can be a crucial decision-making tool for pilots. If a runway was not seen when reaching Decision Altitude (DA), pilots must abort landing.

Configuring the aircraft for landing consisted of several actions that seemed to be procedural. Pilots usually followed manuals and best practice from previous experiences or training. While this action might seem procedural, yet it may involve some ‘improvisations’ depending on the situation. This was usually influenced by pilots’ experience. Pilot 1, for example, during the landing scenario discussed delayed closure of its engine powers when automatic call-outs called ‘retard’ because he could feel the wind pushed the aircraft from above. According to his calculation, if the engine was closed, the aircraft would hit the runway quite hard, thus Pilot 1 decided to delay it for a couple of seconds. This might be the source of high mental demand as pilots put almost their entire focus on outside seeing the runway and PAPI lights while sometime checking the instruments inside.

The last distinct action concluded from the CDM interview was flying and landing the aircraft manually. It was common, as it had been prescribed by manufacturer and company policies, to disconnect the autopilot at a certain distance from the runway. The trigger to start flying manually, however, was different among pilots. Generally, there was a written manual telling pilots when to start flying manually. Yet, the decision may involve personal ‘preference’. Pilot 1, for example, was motivated by gaining back his ability and ‘feeling’ in landing the aircraft after not flying for a while during the Covid-19 pandemic. Therefore, Pilot 1 turned the autopilot off, slightly far from Decision Altitude (DA) given that Pilot 1 could see the runway from the distance. Pilot 5 took a moment after disconnecting the autopilot to check the control of the aircraft. That was the reason Pilot 5 also turned off the autopilot, slightly far from the recommended point (DA).

Workload was not only affected by physical and mental demands, but also performance feedback and internal or external influences. Desired performance feedback may ‘reduce’ workload of pilots. From CDM interviews, we classified three distinct feedbacks of performance. First, feedback was shown by the instrument. Both visual and auditory devices help pilots to know whether their actions were correct or not. For instance, when disconnecting autopilot, a peculiar auditory alarm will be heard by pilots. Visual feedback from the display also told pilots about the results of their actions. Speed, altitude, glide slope, localiser, and direction changes were part of this kind of feedback.

Pilots could also receive feedback from changes in aircraft’s behaviour or environment. When turning the knob for heading changes, for example, pilots could feel and see the aircraft is banking to the right or the left immediately. Regarding the environment, the ability to see the runway during final approach was common performance feedback among pilots. After several changes in plans and being vectored by ATC that make confused, seeing the destination runway could indicate the success of following ATC instructions and executing the plan.

Feedback may also come from the flying partner, as it was certainly part of crew resource management policy. Pilot monitoring, as the name suggests, monitored all essential aspects of the aircraft including the commands from pilot flying. For example, part of configuring the aircraft for landing was lowering the landing gears. While pilot flying asks to lower the gears, pilot monitoring was the one who actually executed the action. After the action was complete, pilot monitoring confirmed to pilot flying telling that 'your request has been done, and it works well'.

Regarding internal or external influences, CDM interviews found two distinct features. First, it came from the 'software' part of the pilots, which consisted of mainly the experience. This includes training and 'stories' from more senior flying partners. The way pilots do decision-making process during a flight was more or less influenced by experience. The story of Pilot 1 delaying power closing due to wind changes could be good evidence. Secondly, it came from the 'hardware' parts, which are manuals and policies. Companies, regulators, and manufacturers have manuals and procedures to support pilots in deciding which actions should be made during the most crucial part of the flight, such as landing. For example, pilots who flew Airbus told that the manufacturer generally recommend to land manually, particularly if landing in the airport with certain ILS category.

The results also suggest that most of the sources mentioned previously were visual. Physical and cognitive task demands, except for ATC communication task, involves mostly visual tasks. The tasks combined activities such as monitoring, planning, and maintaining aircraft's behaviour during manual flying. Performance feedback were also mostly coming from the visual modality, except for confirmation from pilot monitoring, and required pilot to monitor parameters and executed certain actions if needed. From the perspective of pilot flying, tasks involving auditory modality, such as ATC communication, were virtually the secondary task because these tasks 'belonged to' pilot monitoring. These results confirmed previous study (e.g. Wilson, 2002) stating that landing and take-off are a visually demanding task during a flight. CDM interviews provided rich information about the dynamic of landing processes experienced by pilots. However, quantification of these dynamic appears to be unfeasible as the way pilots deal with similar situations might be different. External and internal influences may play an important role in shaping pilots' judgement and decision-making.

4.10 Chapter Summary

This chapter addressed the third research aim of the thesis by investigating situational factors that contribute to pilots' mental workload using CDM interviews to professional pilots. From the study we concluded that the physical and mental demands of a landing processes may come from six distinct actions of pilots, including preparing for the data, communicating with ATC, updating the plan, configuring the aircraft, monitoring outside environment, and landing manually. These possible sources of demands are affected by performance feedback that can be shown visually, auditory, or physically, and influenced by several internal or external factors such as experience, training, 'stories', manuals, procedures, and policies. The findings from this chapter will be used to identify simulated flying tasks within the MATB environment as a 'mental workload generator'. The tasks will be manipulated to demonstrate changes in physiological responses as a proxy to indicate changes in mental workload. The next chapter will provide this demonstration.

Chapter 5

Study 3: Measuring Mental Workload during Simulated Flying Task using Brain Activation and Physiological Indicators

5.1 Chapter Overview

This chapter presents experiments demonstrating how changes in brain activation and physiological indices may indicate changes in mental workload during a simulated flying task. Chapter 5 also aims to address the fourth research aim of this thesis, which is 'to develop a task that can resemble task demands of a flying task and investigate the ability of physiological measures in detecting MWL changes during a flying task'. This chapter comprises two experiments using simulated flying tasks in a MATB environment with various levels of task demands to represent a real flying environment. Experiment 1 explored the ability of fNIRS to respond to changes in mental workload. With several issues raised in Experiment 1, we decided to carry out a further confirmatory experiment with some refinement of the methodology and techniques (Experiment 2). Experiment 2 also broadened the scope of the study by employing additional physiological measures. Another subjective MWL scale was also applied in Experiment 2 to cross-validate the existing NASA-TLX scale.

5.2 Introduction

From Study 2, we identified that pilots experience dynamic interaction between task demands, performance feedback, and external or internal influences in a real flight phase (landing). This interaction may shape different perception of mental workload of a task among pilots. In the landing phase, for example, mental workload may change across the stages within a landing mission, such as approach start point, final approach, and touchdown. These landing tasks are continuous and last for several minutes to hours in real flight, depending on the conditions. The next question to answer is 'could we detect mental workload changes during this kind of dynamic task using brain activation indicator?'. For this purpose, we simulated the cognitive characteristics of a pilot's job in MATB environment. As mentioned in Chapter 2 (Literature Review), MATB can potentially mimic the pilot's job in terms of its simultaneity by providing an environment that imposes cognitive demands on participants (perceptual, motor skill, memory, and decision-making).

This chapter is started with Experiment 1 that aimed to demonstrate that mental workload changes can be detected by changes in brain activation using brain sensors, particularly fNIRS. As mentioned in the literature review chapter, the rationale behind attempts to measure workload using physiological indicators can be put into a simple proposition: "as workload is increased, there is a corresponding increase in the operator's level of arousal (often referred to as an intervening variable) reflected in the activity of the autonomic nervous system" (Sharples & Megaw, 2015, pp. 533). Autonomic Nervous System (ANS) regulates involuntary physiological processes such as heart rate, blood pressure, and digestion. It is composed of two anatomically and functionally different divisions, namely sympathetic nervous system (SNS) and parasympathetic nervous system (PNS). In summary, SNS enables the body to be prepared with stressors via the 'fight-or-flight' response, for example, making blood pressure and heart rate increase. Conversely, PNS promotes the 'rest and digest' processes through, for instance, cardiac relaxation (Waxenbaum et al., 2021).

On a fundamental level, the way the brain works related to changes in mental workload follows the general principle that the level of neuronal activity affects oxygen demands of the brain (Gagnon et al., 2016). When stimulated by certain cognitive activities, neurons in corresponding regions of the brain demand more energy that come from aerobic metabolism of glucose (Fantini et al., 2016). The brain therefore requests more oxygen, and they are transported by blood, resulting in more concentration of oxygenated blood in the activated regions. Regarding brain activation, generally speaking, distinct regions of the brain are responsible for different roles in the body. The most common regional divisions are probably based on the sulcus, or the distinct trench, creating four lobes of the brain: frontal, temporal, occipital, and parietal (Kolb & Whishaw, 2016). The research in this thesis will focus on frontal lobes, particularly prefrontal cortex (PFC), which controls various executive and cognitive functions (Frith & Dolan, 1996). In a more practical context, the prefrontal cortex region is shown ‘activated’ by demanding cognitive tasks such as driving (Foy & Chapman, 2018), air traffic control (Ayaz et al., 2012), or even regulating emotion (Glotzbach et al., 2011).

While one might say that PFC is the central for cognitive functions, it is argued that in many cognitive activities, multiple areas of the brain are activated. The prefrontal cortex, however, is the region that seems to be always activated in a cognitive task (Bell et al., 2006). Our experiment, furthermore, focuses on the exploration concerning the use of sensing technologies in measuring mental workload. This research did not specifically aim to investigate all brain regions that correspond to mental workload changes.

Using the proposition mentioned above, potential techniques to capture oxygenation changes in the brain need to be tested. Functional magnetic resonance imaging (fMRI) has been widely explored for this purpose, mainly due to its ability to produce imaging with high spatial resolution (Mier & Mier, 2015). However, this technique has disadvantages in terms of operational cost, low temporal resolution, and noise. fMRI is also prone to participant motion and inflexible in terms of environment to generate tasks, making it difficult for generating certain functional tasks (Cui et al., 2011).

At this point, fNIRS comes into play. Inspired by a promising demonstration from Jöbsis (1977) to undertake non-invasive measurement of cerebral blood flow, fNIRS now has been utilised to investigate brain activity by detecting changes in oxygenation in particular brain areas (e.g. Ayaz et al., 2012; Pinti et al., 2019). The principle of fNIRS technique involves quantifying haemoglobin concentration resolved from the measurement of near infrared (NIR) light temporal changes. This technique is made possible by haemoglobin characteristic as great light absorbers, along with skin, tissue, and bone that are mainly transparent to near infrared light (with the spectrum around 700-900 nm range). NIR light travels from the emitters through the head to the receivers with ‘banana-shaped’ trajectory. Throughout the journey, the light is scattered or absorbed by the tissue it passes through. Changes in absorbed light can infer changes in haemoglobin concentration, since haemoglobin is a prominent absorber of NIR light. Figure 5.1 shows a graphical illustration of fNIRS technique principle.

The technique is known for its minimum level of obtrusiveness and lower-cost nature,

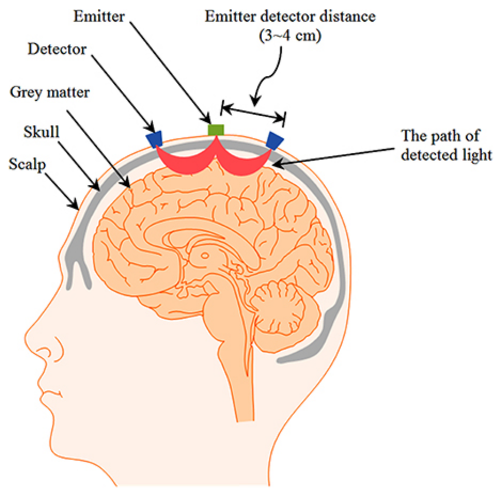


Figure 5.1: The principle of fNIRS technique, taken from Naseer & Hong (2015).

giving investigators flexibility to examine participants in various functional task contexts, such as aviation (e.g. Ayaz et al., 2012; Causse et al., 2017; Durantin et al., 2016; Takeuchi, 2000; Verdière et al., 2018) driving (e.g. Foy et al., 2016; Ahn et al., 2016), or railway (Kojima et al., 2005). Studies using fNIRS have shown that they have close relationship to fMRI signals with high temporal and spatial/linear correlations (Huppert et al., 2006a,b). However, fNIRS still faces challenges related to low spatial resolution and signal-to-noise ratio (SNR), resulting in its limited ability to target only the outer areas of the brain without clear knowledge of how deeply the signals can practically travel inside the brain tissue (Cui et al., 2011). Despite the challenges, fNIRS seems to be a promising technique as it is comparable to fMRI in terms of the validity to measure blood oxygenation changes. Thus, fNIRS can be potentially used to estimate mental workload changes.

5.3 Experiment 1: an fNIRS study

As the main aim of this experiment was to demonstrate whether brain activation changes can be captured by fNIRS, we developed five hypotheses to answer related research questions, as shown in Table 5.1. Regarding hypotheses H5.1.3 and H5.1.4, we added psychological variables: motivation and anxiety. Motivation seems to have relationship with general performance in flying tasks (Frederick-Recascino & Hall, 2003), or with cognitive performance in particular (Smith & Hess, 2015). Besides motivation, anxiety seems to affect performance in flying tasks. Allsop et al. (2016), for example, reported their study on the effect of anxiety on pilot behaviour in scanning instruments. Anxious pilots were reported to have more random gaze behaviour, but only when the cognitive load was high. The inclusion of these variables might inform us whether the role of anxiety and motivation in MWL measurements during multitasks performance matters.

Table 5.1: Research questions and hypotheses for Experiment 1.

No.	Research question	Hypothesis statement
H5.1.1	Will subjective MWL measurement (NASA-TLX) scores show difference between low and high-demand task?	NASA-TLX scores in all dimensions will score higher in high-demand task than low demand task.
H5.1.2	Will performance scores show difference between low demand task and high-demand task?	MATB scores in all tasks will score lower in high-demand task than low demand task.
H5.1.3	Will brain activation in all fNIRS channels show difference between low and high-demand task?	Brain activation in all fNIRS channels will score higher in high-demand task than low demand task.
H5.1.4	Controlling anxiety and motivation, do subjective MWL and performance scores correlate each other?	Controlling state and trait anxiety and motivation, MATB score will negatively correlate with NASA-TLX score.
H5.1.5	Controlling anxiety and motivation, do brain activation and performance scores correlate each other?	Controlling state and trait anxiety and motivation, brain activation will positively correlate with NASA-TLX score.

5.3.1 Methods

Experiment task and design. The design of this study was repeated measures. The independent variable was two different demand levels of MATB tasks: low and high; while the dependent variables were brain activation changes for each channel of fNIRS device (channel 1-8), subjective workload score, and performance score of MATB. The MATB task demands from the work of Kennedy & Parker (2017) were used for this study. Each demand level was defined according to specific setup of the MATB subtasks, particularly regarding the frequency and the difficulty of stimuli. After discussing with the authors regarding MATB-II validation, communication task (COMM) was excluded to avoid differences in stimulus modalities (Lauren-Metz, 2019). Furthermore, from Study 2, it was revealed that in dual-pilot environment handling communication is not the main responsibility of pilot flying. Thus, we focused on visual task demands as this has proven as the main sources of demands during crucial phase of a flight (Wilson, 2002). Moreover, since participants in these experiments were recruited from general rather than specific population with more experience in multitasking abilities (i.e. professional pilots), including auditory stimulus may potentially interfere and impair participants' performance and motivation (Ferraro et al., 2017). Table 5.2 shows the setup details for MATB task demand levels.

Table 5.2: Details of MATB task demand levels from the validation work of Kennedy & Parker (2017).

Level	Tracking (TRACK)	System Monitoring (SYSMON)	Resource Management (RESMAN)
Low demand	Low preset default	two deflections per minute	one pump fails every minute
High-demand	Medium preset default	30 deflections per minute	one to two pumps fail for 15 seconds every minute

In the design of this study, psychological scales (STICSA and MIAMI) were used as control variables in correlation analysis, particularly when testing hypothesis H5.1.4 and H5.1.5.

Participants. Thirty-eight students and staff (M_{age} = 28.00, SD_{age} = 6.48) took part in this experiment. All participants were novices on the specific flying task simulated in this experiment. Participants were recruited mainly from poster advert. If participants agreed to participate, they were asked to read an information sheet and sign a consent before the experiment session. Ethics approval for this study was granted from Faculty of Engineering Research Ethics Committee, the University of Nottingham.

Apparatus There were four apparatus employed for this experiment, as follows:

1. **Artinis OctaMon fNIRS device.** Artinis OctaMon from Artinis Medical System was used to measure changes in haemoglobin activities at prefrontal cortex areas of the brain. Eight channels recorded the activities at a frequency of 10 Hz with two different wavelengths (760 and 850 nm). The device has two near-infrared emitters with four receivers, making eight channels (or optodes) in total, consisting of four channels in peripheral areas of prefrontal cortex (two channels on either sides, approximately corresponds to Brodmann Areas 46) and four channels in central areas of prefrontal cortex (approximately corresponds to Brodmann Areas 9 and 10). Figure 5.2 shows the position of channels.

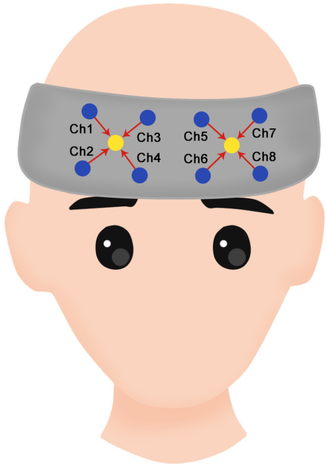


Figure 5.2: Channels location of fNIRS.

2. **Multi-Attribute Task battery.** To induce MWL, the Multi-Attribute Task Battery version 2 (MATB-II) developed by NASA (Comstock & Arnegard, 1992) was employed. MATB was developed to simulate aircraft crew tasks in terms of cognitive demands while not requiring pilot training. The basis for evaluation of performance and workload in MATB comprises four main tasks: (1) system monitoring (SYSMON), which simulates a system monitoring task and requires operator to detect changes in system and respond accordingly; (2) tracking (TRACK), which simulates an aircraft position maintaining task that requires operator to maintain moving target within certain area; (3) resource management (RESMAN), which simulates a fuel management task and requires the participant to maintain fuel level and detect any failures occur in pumping system; and (4) communication (COMM), which simulates communication with air traffic controller (ATC) and requires the operator to detect and respond verbal command from ATC. In addition to the four main tasks, the scheduling (SCHED) window also appears to provide information about expected workload. Figure 5.3 shows the interface of MATB-II.

   MATB was chosen for several reasons. Firstly, MATB has been used in various aviation research (e.g. Bliss, 1997; Caldwell & Ramspott, 1998; Lopez et al., 2012; Wilson et al., 2007; Nixon & Charles, 2017). This appears to be due to its historical relations with aviation research. The tracking task, for example, is theoretically a 'classic' compensatory tracking task that is utilised to test eye-hand coordination and has been used to predict military flight training success since the 1940s (Gibb & Dolgin, 1989). Secondly, MATB has the ability and flexibility to generate multitasking missions. Task scenarios within the software can be easily configured as necessary, giving full control to experimenters for manipulating level of task difficulty or complexity to induce mental workload to participants. Lastly, MATB has been validated by various studies outside aviation context such as driving (Takae et al., 2010), workload or task demands/complexity (e.g. Fairclough et al., 2005; Hsu et al., 2015), multitasking strategy (e.g. Chiappe et al., 2013; Gutzwiller et al., 2016; Nelson et al., 2016), learning environment (Borghini et al., 2016), and cognitive performance (Brannon et al., 2008; Carlozzi et al., 2010).

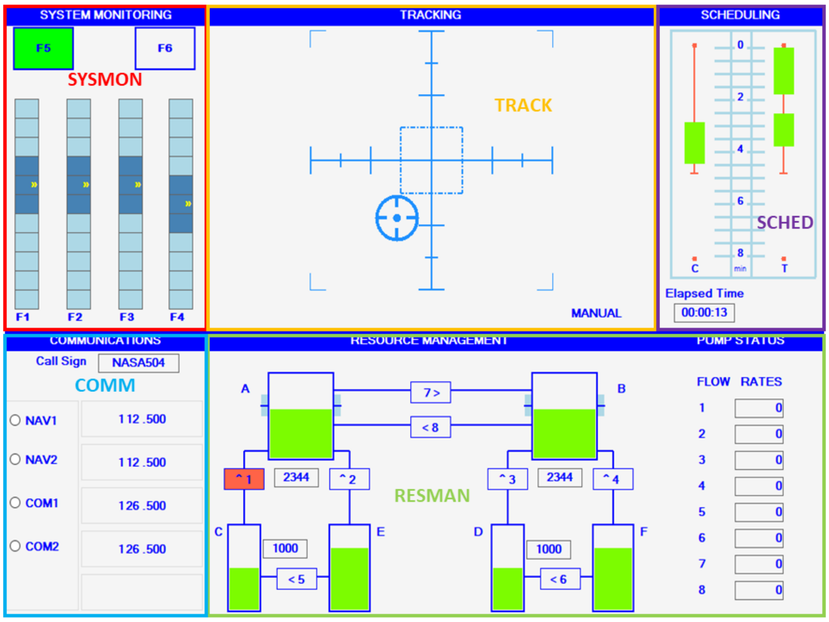


Figure 5.3: MATB-II interface.

In conclusion, MATB may reasonably be said to mimic the real cognitive situation during a flying mission, particularly in the sense of the task's simultaneity. Scores obtained from participants' performance when completing MATB tasks were generated to create performance or behavioural measures.

3. **NASA-TLX.** The scale was administered from MATB software immediately after the completion of each task block. The total raw score of the scales was the main interest for analysis, yet scores from each dimension were also presented. Raw score of the scale was chosen as the main approach to quantify MWL because of its simplicity. Furthermore, it has been concluded that using raw NASA-TLX scale could be more, equally, or less sensitive than the weighted version of the scale (Hart, 2006). Therefore, the justification for using either weighted or unweighted procedures for NASA-TLX scoring appears to be loose, i.e. users may choose the one that fits their study.

4. **Psychological measures**. Momentary Influences, Attitudes, and Motivation Impact (MIAMI) on Cognitive Performance Scale developed by Moritz et al. (2017) was used to assess subjective influences on cognitive performance. The scale has two versions, which are pre/baseline and post. Each version consists of

20 items, covering four domains: (1) poor motivation, (2) concern about the assessment, (3) fear about poor outcome, and (4) negative momentary influences. Participants rated items on a four-point Likert scale (1 for 'fully agree', 2 for 'rather agree', 3 for 'rather disagree', 4 for 'disagree'). The total score for each domain was calculated after adjusting scores for reversed items. Both pre-task and post-task total score was used for analysis. Participants' anxiety was also assessed by using a State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) (Grös et al., 2007). The questionnaire consists of 21 items and participants were required to rate each item ranging from 1-4 according to the degree to which it is true for them (1 for 'not at all', 2 for 'a little', 3 for 'moderately', 4 for 'very much so'). The questionnaire was administered twice, one asking for mood presently (state) and the other asking for mood in general (trait). The total score for both state and trait version of the questionnaire were calculated and used for analysis. Appendix B.4 shows the physical form of these scales.

Procedure. Upon arrival, participants were briefed and given an informed consent form. After declaring their willingness to participate by completing the consent form, participants had a short, ten-minute training on the task, aiming to make them familiar with the interface and the device. Three questionnaires, i.e. MIAMI pre version and STICSA trait and state versions, were required to be completed by participants afterwards. fNIRS device set up and baseline record for brain activity changes were undertaken immediately after all questionnaires completed. Participants were asked to sit and relax while they are closing their eyes for a moment. This procedure also aims to check that the fNIRS device is working properly. A series of tasks consisting of two five-minute easy tasks and two five-minute difficult tasks were employed alternately with two possible starts: easy and difficult start. The selection of low and high-demand start was done by random allocation. NASA-TLX was generated immediately after each task block had been completed. Before the experiment debriefing, participants were asked to complete MIAMI post version questionnaires (see Figure 5.4).

Statistical analysis approach Two-way analysis of variance (ANOVA) was the main statistical approach for testing the differences in variables of interest measures e.g. brain activation, subjective MWL, and physiological response. Demand levels of the MATB task and corresponding factors in each variable of interest, such as fNIRS channels, MATB subtasks, or NASA-TLX dimensions, were considered as factorial design for the ANOVA. Regarding the ANOVA results, we particularly looked for any useful interaction at first, i.e. interaction between demand levels and correspond-

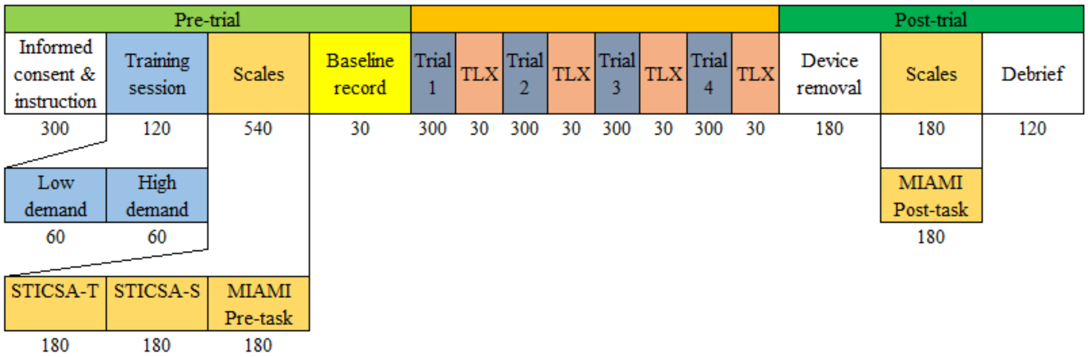


Figure 5.4: Procedure for Experiment 1.

ing factors in affecting changes in variables of interest measures. Mauchly's test was automatically included in the calculation for testing the assumption of sphericity; correction for its violation, if any, was calculated using the Greenhouse-Geisser method. In that case, the corrected values of F-ratio and degrees of freedom were used for reporting the results. If an interaction effect was found, pairwise comparison with Bonferroni correction was performed to locate the differences specifically. Effect size was calculated and interpreted using Cohen's d (Lakens, 2013).

Regarding correlations analysis, i.e. to test hypothesis H5.1.4 and H5.1.5 in particular, we applied three steps of analysis. The first step was to seek the pattern or direction of the correlations, i.e. whether they are mostly positive or negative. The second step was to check whether there were statistically significant correlations between variables of interest. These two steps were undertaken by applying a partial correlation technique with psychological variables (anxiety/STICSA and motivation/MIAMI) acting as control variables. The third step was to test the effect of control variables by examining whether correlation with (first-order correlation) and without (zero-order correlation) control variables were statistically different. All of these analyses were performed in R using 'zeroEQpart' package developed by Richard et al. (2018).

5.3.2 Results

Testing hypothesis H5.1.1: subjective workload scores. Raw total score of the NASA-TLX was used to capture participants' subjective perception towards the task demands. A 2 (low vs. high demand) x 7 (NASA-TLX dimensions and total scores) two-way repeated measure ANOVA with Bonferroni multiple testing correction was performed to evaluate the effect of different dimensions of NASA-TLX tasks and demand levels on participants' subjective rating. There was a statistically significant interaction between demand levels and NASA-TLX dimensions on participants' subjective rating towards the tasks, F(3.36, 124.18) = 3.75, p = 0.01, $\eta_g^2 = 0.008$. Therefore, the effect of demand levels variable was analysed at each dimension of the scale.

Pairwise comparisons, using paired-sample t-tests with Bonferroni correction, indicate that the mean subjective rating of workload was significantly different between low and high-demand task in the following dimensions: mental demand (p = 0.003), temporal demand (p < 0.001), performance (p = 0.033), effort (p < 0.001), frustration (p = 0.001), and total raw score (p < 0.001). However, we did not find significant difference in physical demand (p = 0.100). Moreover, the effect size (using Cohen's d) for these dimensions was found to be none/trivial to small (see Figure 5.5 and Table 5.3 for details).

These results suggest that both demand levels subjectively differ, as captured by all dimensions and total raw scores of NASA-TLX scales. An exception comes from physical demand. Despite showing a similar direction as in other dimensions, i.e. participants tend to rate higher in high-demand task than in low demand task, significant difference was not found. From these subjective results, we may cautiously conclude that task demand manipulation was successful. Participants perceive high-demand task as 'difficult' and vice versa. However, we must note that the differences were not considerable, thus, hypothesis H5.1.1 can be partially supported.

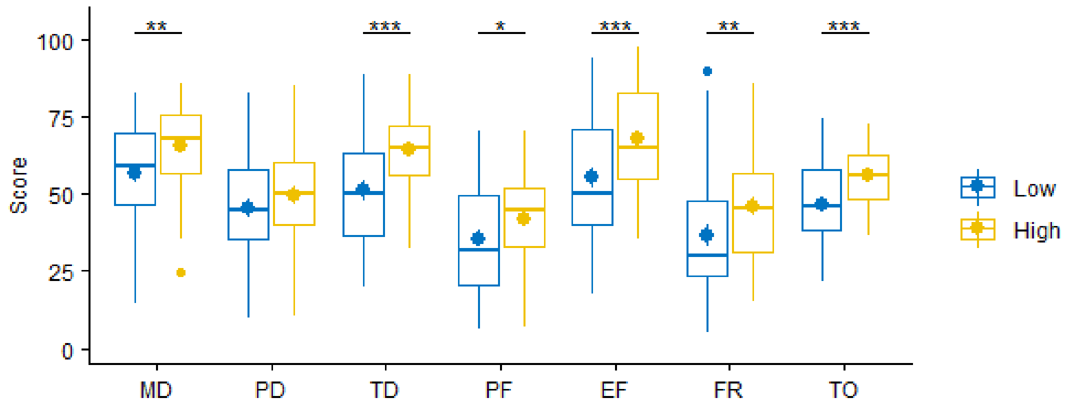


Figure 5.5: Pairwise comparisons of NASA-TLX (Experiment 1). The dots within the box indicate the mean from corresponding NASA-TLX scores, with standard error of the mean. For the total score of MWL, the difference is significant at p < 0.001.

Table 5.3: Mean differences in NASA-TLX (Experiment 1).

Dimension	Level	Mean	SD	SE	p	Cohen's d
MD	Low High	56.724 65.645	17.367 14.236	2.817 2.309	p = 0.003	d = 0.28
PD	Low High	45.447 49.250	17.462 18.539	2.833 3.007	p = 0.100	d = 0.11

(continued on next page)

Table 5.3, continued

Dimension	Level	Mean	SD	SE	p	Cohen's d
TD	Low High	51.118 64.342	19.940 14.313	3.235 2.322	p < 0.001	d = 0.39
PF	Low High	35.487 41.789	17.128 15.591	2.779 2.529	p = 0.033	d = 0.20
EF	Low High	55.184 67.605	21.238 16.203	3.445 2.628	p < 0.001	d = 0.33
FR	Low High	36.526 45.789	20.897 18.740	3.390 3.040	p = 0.001	d = 0.24
TO	Low High	46.750 55.740	14.424 9.550	2.340 1.549	p < 0.001	d = 0.37

Testing hypothesis H5.1.2: MATB performance. Performance or behavioural measures were created from MATB results by averaging normalised scores from each task for each demand levels. As the first step, the strategy used to obtain a score for each task was determined by applying formulas from Kim & Yang (2017). For the system monitoring task, a participant's individual performance was defined as the number of correct responses from the total stimuli introduced during the experiment. Tracking task performance was defined as the root-mean-square deviation from the centre point in pixel units. For the resource management task, performance was defined as the sum of the difference of fuel amount for both tanks from the designated tolerable deviation range (1000 units: 500 above and 500 below zero) during trial samples (recorded every 30 seconds). If the scores reach negative values (performance < 0), it was excluded or considered as zero, as this most probably indicates that participants fail to perform the task most of the time. The scores from the tracking and resource management tasks were inverted (subtracted from 1) to give the impression that higher scores means higher performance. After obtaining these scores, the next step was to normalise the score so that they had a maximum score of 1 and minimum score of 0. The final step of this process was to obtain averaged scores of both low and high demand for each task. By summing scores from these individual tasks, a total performance score for each participant was obtained. Mathematical expressions of these performance score definitions can be seen in Appendix A.1.

Using the scores, a 2 (low vs. high demand) x 4 (SYSMON, TRACK, RESMAN, TOTAL) two-way repeated measures ANOVA with Bonferroni multiple testing correction method was performed to evaluate the effect of different MATB tasks and demands on performance. There was a statistically significant interaction between demand levels and MATB tasks on performance score, $F(1.89, 69.76) = 27.34$, $p < 0.0001$, $\eta^2_g = 0.05$. Therefore, the effect of demand levels variable was analysed for each

MATB task. Pairwise comparison, using paired-sample t-tests with Bonferroni correction, show that the mean performance score was significantly different between low and high-demand task in system monitoring task ($p < 0.001$), tracking task ($p < 0.001$), and all tasks together (TOTAL) ($p < 0.001$), but not in resource management task ($p = 0.22$). The effect size also varies for these factors, from none/trivial (resource management), small (system monitoring), medium (total performance), and large (tracking) (see Figure 5.6 and Table 5.4).

These results suggest that both demand levels practically differ, except in the resource management task, even though the difference in scores were similar to other tasks (higher in low demand task, vice versa). From this performance (behavioural) scores, we may conclude that high-demand task results in poorer performance score because the task generates considerably more demands than low demand task. Hypothesis H5.1.2 therefore can be partially supported.

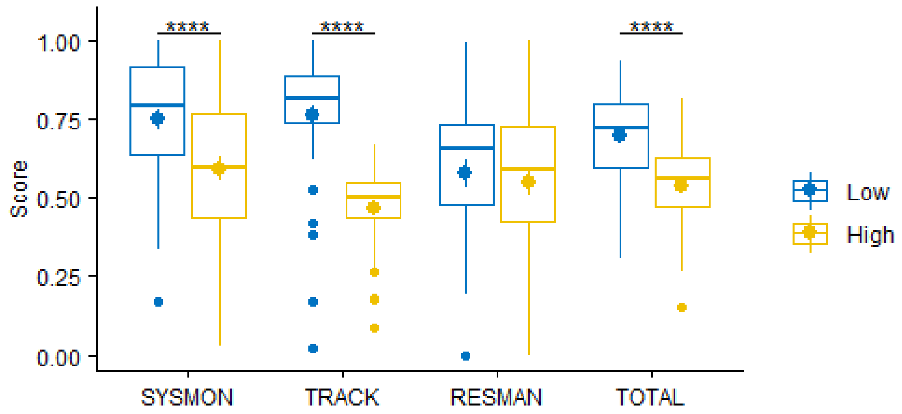


Figure 5.6: Pairwise comparisons of MATB performance scores (Experiment 1). The dots within the box indicate the mean from corresponding MATB task scores, with standard error of the mean. For the MATB total score, the difference is significant at $p < 0.0001$.

Table 5.4: Mean differences in MATB performance scores (Experiment 1).

Task	Level	Mean	SD	SE	p	Cohen's d
SYSMON	Low High	0.747 0.591	0.215 0.232	0.035 0.038	p < 0.0001	d = 0.35
TRACK	Low High	0.759 0.467	0.211 0.135	0.034 0.022	p < 0.0001	d = 0.83
RESMAN	Low High	0.575 0.550	0.264 0.260	0.043 0.042	p = 0.22	d = 0.04
TOTAL	Low High	0.694 0.536	0.139 0.139	0.023 0.022	p < 0.0001	d = 0.57

Testing hypothesis H5.1.3: brain activation. The fNIRS data from participants was live-streamed via Bluetooth signals to Oxysoft software and computer supplied by the company. Since raw fNIRS data cannot be directly used for analysis due to its confounding with unwanted signals or noise, data pre-processing is essential. Homer2 software (Huppert et al., 2009), a graphical user interface (GUI) programme run in Matlab (Mathworks Inc., Sherborn, MA, USA), was employed for this task. However, there are no universal guidelines on how to pre-process fNIRS data as it may depend on the characteristics of the studies, participants, or devices. We therefore set a preprocessing pipeline for this study as seen in Table 5.5. This was combined from previous research showing successful attempts in correcting the signal.

Table 5.5: Preprocessing pipeline for fNIRS data analysis.

Processing steps	**Description**	**Function in Homer2**	**Parameters**
Intensity to optical density	Converting raw data to optical density	hmrIntensity2OD	n/a
Channel pruning	Removing channels when signals were too weak or too strong	enPruneChannels	SNR threshold = 2; dRange = 1e-2 to 3e (Wolff et al., 2019)
Wavelet filtering	Correcting motion artefacts by applying discrete wavelet transform	hmrMotionCorrect-Wavelet	α = 0.1 (Molavi & Dumont, 2012); or interquartile range (IQR) = 1.5 (Behrendt et al., 2018)
Motion artefacts removal	Removing motion artefacts by identifying signal change greater than designated threshold	hmrMotionArtifact	tMotion = 0.5; tMask = 2.0; STDEVthresh = 20.0; AMPthresh = 0.5 (Cooper et al., 2012)
Bandpass filter	Removing instrument and physiological noises	hmrBandpassFilt	low-pass = 0.5 Hz; high-pass = 0.0 Hz (Pinti et al., 2019)
Optical density to concentration	Converting optical density to haemoglobin concentration	hmrOD2Conc	default differential path length factor (DPF) = 6.0 for both frequencies (Scholkmann & Wolf, 2013)
Block average	Averaging haemodynamic response function (HRF) within a block of trial	hmrBlockAvg	baseline correction = -30s; time range of trial = 300s

Figure 5.7 shows the comparison of original (shown in black line) and corrected fNIRS signals (shown in red line; oxygenated haemoglobin only). From the figure, it can be seen that corrected signals (the lines) appeared to be smoother and close to the original one. Based on this image, we interpreted that the corrected signals quality was satisfactory and free from artefacts.

The data from several participants needed to be removed as they did not pass predefined correction and filtering criteria. More specifically, they were excluded as their raw signals exceeded the channel pruning parameter (SNR threshold=2; dRange=1e-2 to 3e). These are listed in Table 5.6. Figure 5.8 shows the flow of processing steps with the number of participants being rejected, if any, on each step.

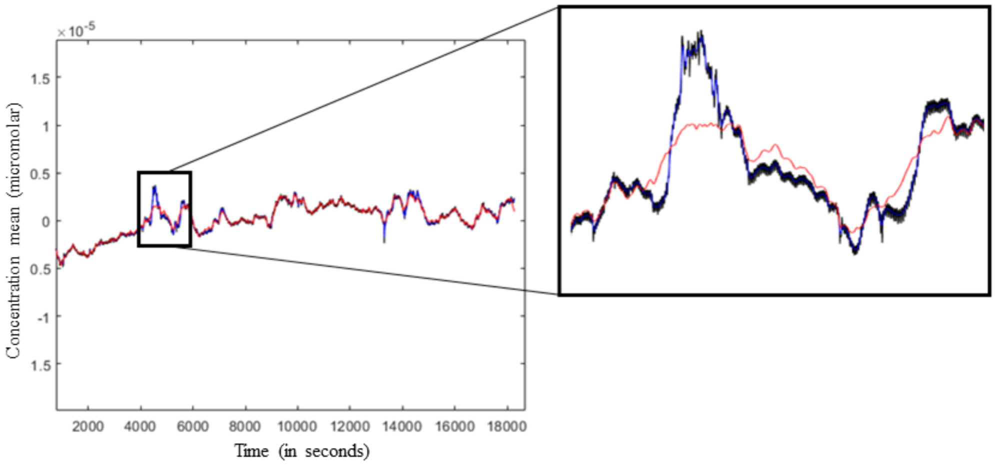


Figure 5.7: Corrected signals after artefacts removal using pipeline.

Table 5.6: Participants whose data was excluded in Experiment 1.

Channel	Participant's ID
Ch.1	6, 16, 27
Ch.2	11, 16
Ch.3	7, 8
Ch.4	n/a
Ch.5	20, 38
Ch.6	24
Ch.7	1, 19, 23
Ch.8	23

To deal with individual variations in fNIRS response, we first normalised all individual oxygenated haemoglobin (HbO) and deoxygenated haemoglobin (HbR) values by dividing them with the difference between maximum and minimum values for each channel (Sato et al., 2005). Since activation is typically defined as an increase in HbO and a slight decrease in HbR, we then produced an index to estimate brain activation based on the differences in both types of haemoglobin for all channels as suggested by Ayaz et al. (2012) i.e. by subtracting the concentration of deoxygenated haemoglobin from the concentration of the oxygenated haemoglobin; both are expressed in micromolar unit (see Appendix A.2 for the mathematical expression). Therefore, the more positive value of this index indicate the more brain activation and vice versa.

We then tested whether there was a significant increase in brain activity during a task block in each channel by employing a one-sample t-test against zero (Lu et al., 2015). The results show that activation occurred in three channels when performing

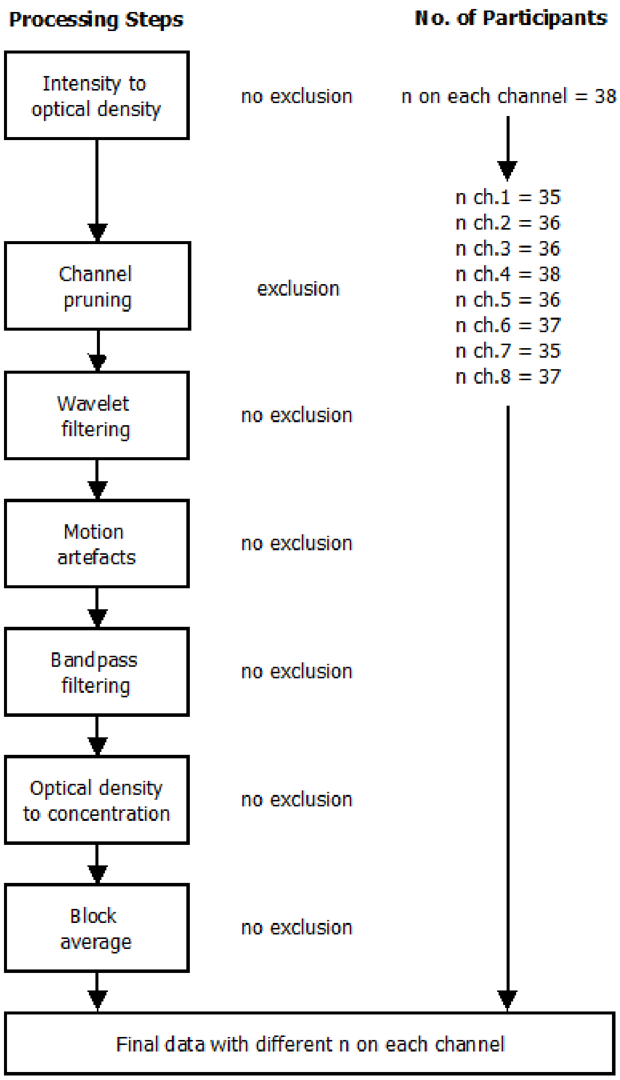


Figure 5.8: Data exclusion flow with the number of included participants for Experiment 1.

low demand task and also three channels when performing high-demand task. Table 5.7 shows the details of the activated channels.

Table 5.7: One-sample t-test results for activated channels (Experiment 1).

Level	Channel	Mean	SD	t statistics	p
Low	Ch.2	0.091	0.232	t(35) = 2.361	0.024
	Ch.4	0.110	0.313	t(37) = 2.169	0.037
	Ch.7	0.110	0.313	t(34) = 2.432	0.020
High	Ch.1	0.123	0.276	t(34) = 2.638	0.013
	Ch.2	0.096	0.275	t(35) = 2.090	0.044
	Ch.7	0.100	0.285	t(34) = 2.070	0.046

A 2 (low vs. high demand) x 8 (Channel 1 to 8) repeated measure ANOVA was used to compare oxygenation changes (HbDiff) of participants during two demand levels of simulated flying tasks (MATB). We did find significant interaction between task demand levels and fNIRS channels on brain activation (F(4.16, 108.08) = 3.26, p =

0.013, $\eta_g^2 = 0.005$). However, after performing pairwise comparison using paired-sample t-test with Bonferroni correction, we did not find any significant differences in all channels. The results may suggest that in this experiment, the difference in task demands appears to be insufficient to trigger significant brain activation in all channels. As seen in Figure 5.9, brain activation scored higher in high-demand task than in low demand task, with exceptions for channels 4, 5, and 7. Nevertheless, the mean differences between these two demand levels were negligible, with trivial effect size was found in all channels (see Table 5.8). Based on these results, we therefore do not have enough evidence to support hypothesis H5.1.3.

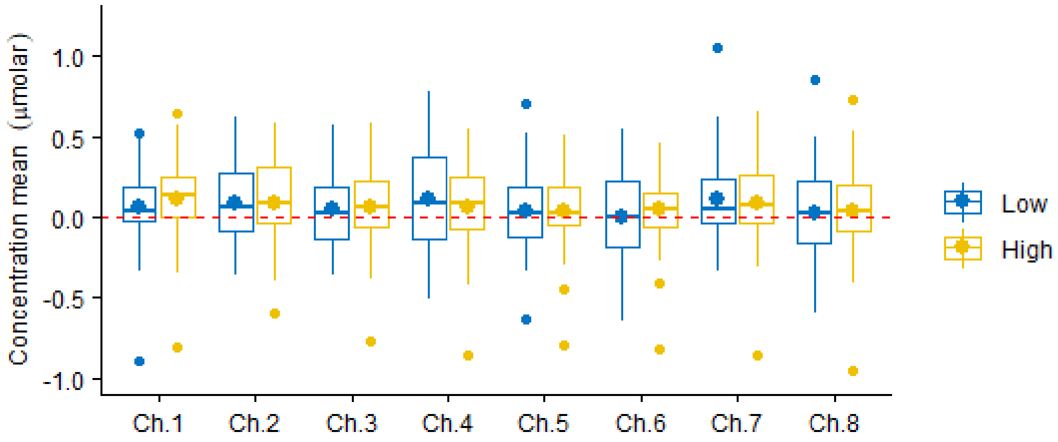


Figure 5.9: Comparisons of brain activation between channels (Experiment 1). The dots within the box indicate the mean from brain activation index with standard error of the mean. The more positive values indicate the more brain activation and vice versa.

Table 5.8: Mean differences in brain activation scores (Experiment 1).

Channel	Level	Mean	SD	SE	p	Cohen's d
Ch.1	Low High	0.063 0.123	0.264 0.276	0.045 0.047	p = 0.176	d = 0.11
Ch.2	Low High	0.091 0.096	0.232 0.275	0.039 0.046	p = 0.923	d = 0.01
Ch.3	Low High	0.056 0.067	0.235 0.269	0.039 0.045	p = 0.810	d = 0.02
Ch.4	Low High	0.110 0.066	0.313 0.260	0.051 0.042	p = 0.377	d = 0.08
Ch.5	Low High	0.036 0.035	0.262 0.258	0.044 0.043	p = 0.988	d = 0.00
Ch.6	Low High	0.005 0.047	0.278 0.252	0.046 0.041	p = 0.406	d = 0.08
Ch.7	Low High	0.116 0.100	0.281 0.285	0.048 0.048	p = 0.769	d = 0.03
Ch.8	Low High	0.029 0.038	0.298 0.287	0.049 0.047	p = 0.859	d = 0.01

Testing hypothesis H5.1.4: MATB-TLX correlation. For testing this hypothesis, a partial correlation was run to determine the relationship between performance scores and subjective MWL scores whilst controlling for anxiety and motivation. There was a small, negative partial correlation between subjective MWL scores with system monitoring task ($r(71) = -0.217$, $p = 0.065$) and resource management task ($r(71) = -0.250$, $p = 0.033$), and a moderate, negative correlation with tracking task ($r(71) = -0.379$, $p = 0.001$) and total MATB performance score ($r(71) = -0.415$, $p < 0.001$). Moreover, significant correlations were also found for all of these relationships, except with system monitoring. Partial correlation analysis also showed that difference between zero- and first-order correlation was found to be high in system monitoring task and total MATB performance, whilst not much different in tracking and resource management task. This may suggest that anxiety and motivation had a certain degree of influence in controlling for the relationship of MWL-MATB total score and MWL-SYSMON. However, the influence tended to be very little in controlling MWL-TRACK and MWL-RESMAN relationships. From the results, therefore, hypothesis H5.1.4 can be partially supported. Table 5.9 summarises the results of this correlation analysis.

Table 5.9: Correlation analyses summary for MWL-MATB (Experiment 1).

Var.	Corr.	SYSMON	TRACK	RESMAN	TOTAL
MWL	Zero-order	-0.006	-0.384**	-0.263*	-0.331**
MWL	First-order	-0.217	-0.379**	-0.250*	-0.415***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Testing hypothesis H5.1.5: brain activation-TLX correlation. A partial correlation was run to determine the relationship between brain activation and subjective MWL scores whilst controlling for anxiety and motivation. There was a small, positive partial correlation between subjective MWL scores from NASA-TLX and all fNIRS channels (see Table 5.10). However, we did not find any significant correlation for all of these relationships. From the partial correlation analysis as well, zero-order correlation did not show difference with first-order correlation in terms of the pattern/direction and significance level, indicating that anxiety and motivation has very little influence in controlling for the relationship between MWL and brain activation. Table 5.10 summarises the results of this correlation analysis.

Table 5.10: Correlation analyses for MWL-brain activation (Experiment 1).

Var.	Corr.	Ch.1	Ch.2	Ch.3	Ch.4	Ch.5	Ch.6	Ch.7	Ch.8
MWL	Zero-order	0.108	0.143	0.129	0.087	0.106	0.125	0.068	0.140
MWL	First-order	0.079	0.121	0.097	0.107	0.110	0.125	0.071	0.123

From the results, the direction of the relationship was found to be as hypothesised. However, since we did not find any significant correlation, hypothesis H5.1.5 thus cannot be supported.

5.3.3 Discussion

This experiment aimed to investigate whether mental workload changes during a simulated flying task can be detected by changes in brain activation using fNIRS. Before further discussion, we first revisit all hypotheses statements and testing results as shown in Table 5.11.

Table 5.11: Hypotheses testing results for Experiment 1.

No.	Hypothesis (alternative) statement	Hypothesis testing
H5.1.1	NASA-TLX scores in all dimensions will score higher in high-demand task than low demand task.	Partially supported
H5.1.2	MATB scores in all tasks will score lower in high-demand task than low demand task.	Partially supported
H5.1.3	Brain activation in all fNIRS channels will score higher in high-demand task than low demand task.	Not supported
H5.1.4	Controlling state and trait anxiety and motivation, MATB score will negatively correlate with NASA-TLX score.	Partially supported
H5.1.5	Controlling state and trait anxiety and motivation, brain activation will positively correlate with NASA-TLX score.	Not supported

Hypothesis H5.1.1 and H5.1.2 aimed to check manipulation of demands in MATB tasks. With different levels of demand, high task demand will practically generate high mental workload and vice verse. Changes in task demand will then theoretically be reflected in participants' perception of the task as self-reported in NASA-TLX scale and performance of the task itself. Participants were expected to report high mental workload and score lower in their performance when performing high demand MATB task. This applied oppositely when performing low demand MATB task.

The results show that manipulation of the task worked, indicated by perceived workload score and performance score. However, not all elements of both variables shows significant different between the two levels of demand. In NASA-TLX scale, the difference in physical demand was not statistically significant. Participants might consider that both levels of demand did not differently exercise their physical effort. This might be due to the nature of the tasks that were mostly visual and the only direct physical activity involved was moving the joystick or pressing the keyboard. The rest of NASA-TLX dimensions, moreover, tend to have no relation to more tangible and direct physical experience as in the physical demand (PD). While participants might be able to directly assess their physical effort as the basis for evaluating PD, connecting dimensions such as the mental demand (MD), the temporal demand (TD), the performance (PF), the frustration (FR), and the effort (EF) to more physical experience appears to be unlikely. Thus, in this case, participants might have different basis of evaluating PD and non-PD dimensions of NASA-TLX, yielding different range of perceiving low or high-demand task. Despite non-significant difference of PD, the total raw score of NASA-TLX showed significant difference in perceiving workload of high and low demand task. We might conclude therefore that the task successfully generated different mental workload levels.

This conclusion was supported by the results from performance scores analysis. The total performance scores of MATB showed significant difference between high and low demand. This might indicate that the workload levels the task generated were practically different. While NASA-TLX was subjective, performance score may be considered an 'objective' indicator to the task 'difficulty'. However, not all subtasks of MATB showed significant difference. Participants scored almost indifferent between the two demand levels when performing the resource management task. This might be due to the nature of resource management task that seems less dynamic than system monitoring and tracking tasks. In both tasks, changes in the objects of interest tend to be fast, thus requiring immediate responses. Meanwhile, in the resource management task, the participants' job was mainly monitoring the tanks as the fuel levels gradually drained and applying strategies to balance them. The nature of this task thus tended to be passive and can be temporarily ignored, since the act of balancing the tanks did not occur 'all the time'. Since performance score agreed with subjective evaluation of the task workload, we may confidently conclude that the task had generated different levels of mental workload.

Despite differences in task demand levels thus mental workload as shown by performance and subjective workload scores, we have no evidence from this experiment to show that fNIRS can detect changes in mental workload. The results indicated that brain activation had occurred in Channel 2, 4, and 7 in low demand task, and in Channel 1, 2, and 7 in high-demand task. This activation may indicate that the task have triggered cognitive activity within the brain function. Theoretically, the brain subsequently requested oxygen for producing energy, thus higher concentration of oxygenated blood could be observed in the region of interest. The results from this experiment furthermore showed the expected pattern; oxygenated blood concentration was found to be slightly higher in high-demand task in most channels (except Channel 4 and 5). However, the concentration differences between task demand levels were not statistically significant. With the results from fNIRS data, we may conclude that the task could trigger the brain to be activated on the region of interest yet fNIRS is seemingly insufficiently sensitive to detect the differences.

The results from Experiment 1 led to an evaluation of the design of the experiment and resulted in identifying three potential refinements for the experiment in the future. The first issue concerned insufficient resting period. In our experiment, we had four trials that went continuously with slight interruption of administering NASA-TLX after the task. We suspected that proceeding to the next trial after NASA-TLX administration without providing a new baseline recording might have ruined the measurement. It is suggested that a resting period after completion of each task block should be added to let brain cortical activity return to baseline level after a task. The design of the task in this experiment aimed to resemble real-world flying tasks in terms of its cognitive demands and continuity. In this case, we put all trials next to each other directly (without resting in between) and NASA-TLX administration was considered part of the task. However, since fNIRS measures relies on changes in oxygenation levels between baseline and tasks, our task design appears to be inadequate. Moreover, the design might not be ideal to compare the data between trials. Therefore, task design changes have to include baseline/resting period recording after NASA-TLX administration in each trial. With the revised design, the data could be more comparable.

However, ideal duration for resting period seems to vary depending on the characteristics of the study and trade-off between random physiological fluctuations and task-unrelated activation (mind-wandering), from as short as two seconds to 600 seconds at the longest (Herold et al., 2018). In Experiment 1, we applied 30-second baseline recording, as suggested by our fNIRS device manual and training. Since our task can be considered ‘long enough’ (five minutes) for fNIRS recording, extension of resting period appears to be essential. The one-minute baseline period was apparently ideal for task design similar to ours. According to our justification, this could be a middle point between allowing oxygenation level to return to its baseline but without triggering mind-wandering due to a long period of ‘doing nothing’.

The second issue that needs to be addressed was participants’ familiarity with the task. We presumed that familiarity with the tasks was not sufficient as the result of poor training session before the experiment. In this particular study, participants had approximately five-minutes training on MATB task using preset scenarios from MATB. However, the task during the training session was different to the task in the experiment session in terms of duration and demand. After evaluating it, the task during training session tended to be ‘low demand’. This may cause participants to perceive the first trial in the experiment session, regardless of the demand levels, as ‘difficult’ because the previous one might have been perceived as ‘easy’. We have attempted to look at the data according to the task sequences (low demand vs high-demand start) to confirm this suspicion. Data from oxygenation concentration levels, MATB performance, and subjective workload rating confirmed this suspicion. Participants from the ‘low demand start’ group scored similar to their ‘high-demand start’ group counterparts in the first trial. The similarity also appeared in the following trials, regardless of the demand levels of the task (See Figure 7.2). Therefore, we assumed that the training session using task scenarios with similar characteristics to the experiment session (in terms of duration and demands) needs to be provided. This strategy aimed to make sure that participants understand the task thoroughly.

The third issue was about confirming fNIRS measurements with other physiological devices. From the results, we cannot be confident to conclude that fNIRS was not sensitive for the task. Other physiological indicators of mental workload that have been theoretically tested are needed to confirm this. If, for example, all physiological measurements cannot show differences between low and high-demand task, we might suggest that the task is not sensitive for triggering physiological response. In other words, these physiological measurements are not fit for detecting changes in MWL during that kind of task. For this reason, adding other physiological measurements such as heart rate or eye-tracker would be beneficial for generating robust conclusion regarding this matter. Not only physiological measures, subjective measurement of MWL needs to be confirmed by other subjective measurements. Since NASA-TLX is subjective, adding other measurements could provide confidence in explaining task manipulation success. Moreover, NASA-TLX was administered after a long task, thus prone to misjudgment. However, administering NASA-TLX in the middle of the task was not possible, as it could interrupt the task. Therefore, immediate measure of subjective workload, such as Instantaneous Self-Assessment of Workload (ISA) (Brennan, 1992), could be a good candidate to cross-validate NASA-TLX measures.

Regarding correlation analyses (hypothesis H5.1.4 and H5.1.5), we found that the direction and magnitude of the correlation was low to medium, with statistical significance found in total, tracking, and resource management tasks of MATB. This result may suggest that performance of the MATB task can be used as an indicator of mental workload, as it correlates with perceived mental workload measure. MATB task that generates high workload tends to result in low performance score by participants. Collaterally, they also perceive the task as 'high-demand task' as captured by NASA-TLX. From this point, performance score can explain mental workload generated by a task. Psychological condition of participants may contribute to this relationship. While the difference in tracking and resource management tasks appears to be negligible, the correlation of subjective MWL and total score of MATB changes from -0.331 to -0.415; and of subjective MWL with system monitoring task from -0.006 to -0.217. This may suggest that performance in MATB task is seemingly dependent on psychological conditions (anxiety and motivation) of participants, particularly when performing system monitoring task.

Concerning fNIRS, we did not find any significant correlation between subjective MWL score and all fNIRS channels. However, the direction of the relationships was as expected, suggesting that increase in oxygenation concentration in the brain was followed by increased perception towards workload of the task. The magnitude of the relationships were found to be small, ranging from 0.068 to 0.143. With the data showing not many differences in oxygenation levels between low and high-demand task, this result can be expected. The difference between zero- and first-order correlation also did not show much different, suggesting brain activation during the MATB task was not influenced by participants' anxiety and motivation. These correlation results also corroborate our suspicion about the design of Experiment 1 that appears to be insensitive for such MATB task. Hence, a confirmatory experiment with refined task design needs to be undertaken to test the notion.

The next experiment replicated this experiment with several corrections previously discussed.

5.4 Experiment 2: fNIRS and physiological studies

Theoretically, Experiment 2 was identical to Experiment 1 with some refinement of methodology and techniques as mentioned in the discussion section of Experiment 1. This experiment aimed to verify the results of the fNIRS method in detecting MWL. It also aimed to broaden the scope of MWL changes detection by applying physiological measurements such as heart rate and pupil dilation and analysing more correlations between variables.

Regarding heart rate, it has been mentioned previously that it transports blood as requested by the brain for energy to process cognitive functions. These chain processes may also explain the relationship between the brain and the heart. It is argued that there is a pathway from the brain to regulate heart rate via both sympathetic and parasympathetic nerves (Thayer & Lane, 2009). In simpler terms, brain activation will be most likely followed by heart rate changes to adjust cerebral blood volume. Specifically for mental workload estimation, HR and HRV measures act oppositely, i.e. HRV measures tend to decrease following the increase in mental activities and vice versa (Delliaux et al., 2019). There are mainly two indices of HRV, namely time-domain and frequency-domain. Time-domain indices of HRV use the degree of variability in measures of the inter-beat interval (IBI), whereas frequency-domain measurements use the absolute or relative power spectral density distribution as a function of frequency (Shaffer & Ginsberg, 2017). For this experiment, we decided to use time-domain analysis (SDNN and RMSSD) as our data recording was shorter than 24 hours. Time-domain analysis is also considered comparable to frequency-domain, and easier to be performed (Malik et al., 1996).

In addition to heart rate, pupil diameter is thought to be an indirect cue of the brain state. Changes in pupil diameter are controlled by two antagonistic pupillary muscles: the dilator pupillae which dilates the pupil and the sphincter pupillae which constricts it. Both pupillary muscles receive inputs from brain parts that are responsible for cognitive and autonomic functions through the activation of both sympathetic and parasympathetic systems (Eckstein et al., 2017). Consequently, changes in pupil size may serve as a proxy for brain activation triggered by mental demands (Menekse Dalveren et al., 2018).

For the purposes previously mentioned, we set 11 hypotheses for this experiment as in Table 5.12

Table 5.12: Research questions and hypotheses for Experiment 2.

No.	Research question	Hypothesis statement
H5.2.1	Will subjective MWL measurement (NASA-TLX) scores show difference between low and high-demand task?	NASA-TLX scores in all dimensions will score higher in high-demand task than low demand task.
H5.2.2	Will instantaneous subjective MWL measurement (ISA) scores show difference between low and high-demand task?	ISA scores in all dimensions will score higher in high-demand task than low demand task.
H5.2.3	Will NASA-TLX correlate with ISA?	NASA-TLX will positively correlate with ISA.
H5.2.4	Will performance scores show difference between low demand task and high-demand task?	MATB scores in all tasks will score lower in high-demand task than low demand task.

(continued on next page)

Table 5.12, continued

No.	Research question	Hypothesis statement
H5.2.5	Will brain activation in all fNIRS channels show difference between low and high-demand task?	Brain activation in all fNIRS channels will score higher in high-demand task than low demand task.
H5.2.6	Will heart rate and its variability show difference between low and high-demand task?	Heart rate will score higher in high-demand task than low demand task, whilst its variability measures will score oppositely.
H5.2.7	Will pupil diameter show difference between low and high-demand task?	Pupil diameter in both eyes will dilate wider in high-demand task than low demand task.
H5.2.8	Controlling anxiety and motivation, do subjective MWL and performance scores correlate each other?	Controlling state and trait anxiety and motivation, MATB score will negatively correlate with NASA-TLX score.
H5.2.9	Controlling anxiety and motivation, do subjective MWL and brain activation scores correlate each other?	Controlling state and trait anxiety and motivation, brain activation will positively correlate with NASA-TLX score.
H5.2.10	Controlling anxiety and motivation, do subjective MWL and HR/HRV scores correlate each other?	Controlling state and trait anxiety and motivation, HR will positively correlate with NASA-TLX score; whilst RMSSD and SDNN will negatively correlate with NASA-TLX score.
H5.2.11	Controlling anxiety and motivation, do subjective MWL and pupil diameter correlate each other?	Controlling state and trait anxiety and motivation, pupil diameter will positively correlate with NASA-TLX score.

Figure 5.10 shows relations between variables of interest and hypotheses in Experiment 2.

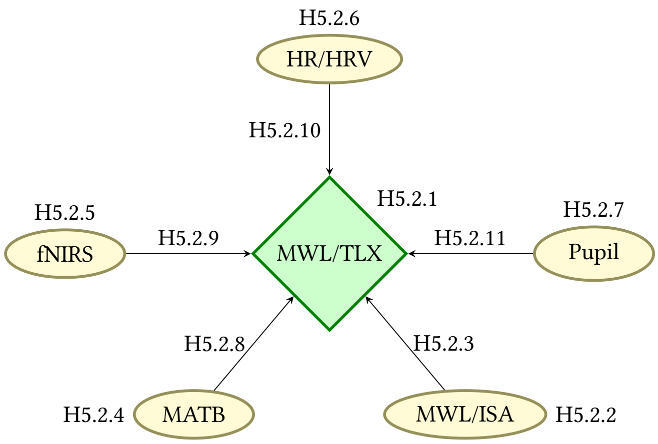


Figure 5.10: Relations between variables and hypotheses in Experiment 2.

5.4.1 Methods

Experiment task and design. A within-subject design as in Experiment 1 was used, with task demand levels (low and high) served as independent variable. The dependent variables were brain activation changes for each channel of fNIRS device (Ch.1-8), subjective workload score (NASA-TLX and ISA), performance score of MATB, with heart rate measures and pupil diameter as the additional dependent variables. MATB was used for this experiment and the tasks had similar properties as in Experiment 1. The differences were that we rearranged the events within each task for each demand level using a Matlab script; thus their arrangements would be completely random. Furthermore, we programmed all possible configurations for sequences of the tasks, resulting six different sequences, namely ABAB, AABB, ABBA, BABA, BBAA, and BAAB (A is indicating 'low' and B is 'high'). This was different to Experiment 1 that had only two configurations (ABAB and BABA). Participants threw a digital dice before the experiment session to determine their configuration.

Participants. Thirty participants mainly recruited from university students and staff (M_{age} = 32.04, SD_{age} = 4.74) took part in this experiment. All participants were novices on the specific flying task simulated in this experiment. Participants were invited mainly from a poster advert. If participants agreed to participate, they were asked to read an information sheet and sign a consent before the experiment session. Ethics approval for this study was granted from Faculty of Engineering Research Ethics Committee, the University of Nottingham.

Apparatus. Apparatus were identical to Experiment 1 (see Section 5.3.1 p.110), including the fNIRS device and the preprocessing pipeline. Moreover, we added three apparatus, as follows:

1. **Heart rate measures.** Zephyr Bioharness 3 made by Zephyr Technology Corporation (Annapolis, MD, US) was used to capture heart rate thus heart rate variability (HRV). It is a physiological monitoring device that can record heart activities and breathing rate. The device was attached to a chest strap equipped with both heart rate and breathing sensors (Zephyr Technology, 2012). The data was directly saved in the device and can be transferred later to a computer for analysis. Figure 5.11 shows the illustration of Zephyr Bioharness when being worn by a person.

Figure 5.11: Zephyr Bioharness 3 Sensors.

2. **Eye-tracker.** SMI Eye Tracking Glasses from iMotions A/S (København K, Denmark) were utilised in this experiment, along with temporary availability of Tobii Pro Eye Tracking Glasses 2 from the same manufacturer. However, we mostly utilised SMI Eye Tracking Glasses for this experiment purpose. These two eye-tracking devices were developed to be used with adult participants specifically for the research context. The head unit, similar to a common pair of glasses, must be worn onto the test participant's head to collect eye-tracking data. The system must be independently calibrated for each participant, by asking them to look at a special calibration card for a few seconds. The glasses must be cable-connected to a computer running the software. This software is a command centre for controlling all interactions with the glasses from calibration, experiment or test run, recording start/stop, etc. During a test or experiment session, the researcher can hear and see what is being recorded by the glasses, including participants' gaze point shown in a coloured dot. Figure 5.12 shows the application of the glasses along with another sensor (fNIRS on the forehead).

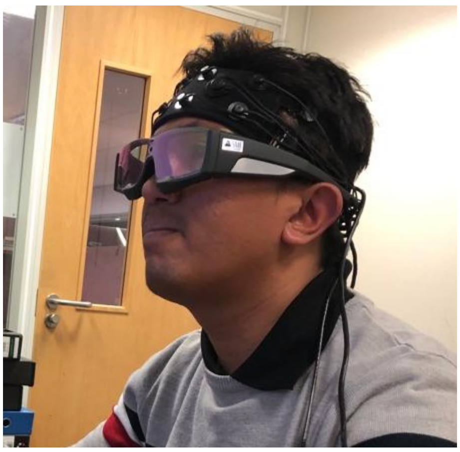


Figure 5.12: A participant wearing SMI eye-tracking glasses.

3. **ISA workload scale.** To validate and complement subjective workload mea-

sures from NASA-TLX, we also employed Instantaneous Self-Assessment of Workload (ISA) from Brennan (1992). ISA, as the name suggests, provides immediate subjective evaluation of mental workload during a task and was originally developed for assessing air traffic controllers (ATC) workload. Its instantaneity makes this scale less intrusive to the task, thus capable of real-time assessment. We have presented a brief description of this scale in Chapter 2 (Literature Review); Appendix B.4 shows the guideline for using ISA.

Procedure. Upon arrival, participants had a short explanation about the experiment and, if they agreed to participate, an informed consent form was completed and signed. The participant had a 10-minute training session, asking them to experience the MATB task for both task demand levels (low and high), including ISA and TLX administration. Immediately after the training session ended, participants were asked to complete STICSA trait version, STICSA state version, and MIAMI pre-task scales. Device setup was undertaken after completing the scales. Each task block was started by asking participants to sit and relax for a one-minute baseline measurement while staring to the monitor (resting condition). A one-minute baseline was chosen as we considered this as the optimum 'intersection' between having an overly short baseline (i.e. 30 seconds, as in Experiment 1) that appears to be suboptimal for this kind of experiment, and having an overly long baseline that will produce unwanted brain activation due to mind-wandering (Herold et al., 2018). The first task block was started immediately, and it lasted for five minutes, followed by one-minute rest/baseline again and thirty-second NASA-TLX administration. The second to fourth task blocks followed with the same protocol. As mentioned earlier, the order of the block (low or high-demand start) was randomly determined by throwing a digital dice. After completing four task blocks, all devices were removed. Participants left with their monetary compensation after the MIAMI post-task had been completed. Figure 5.13 shows the experiment procedure.

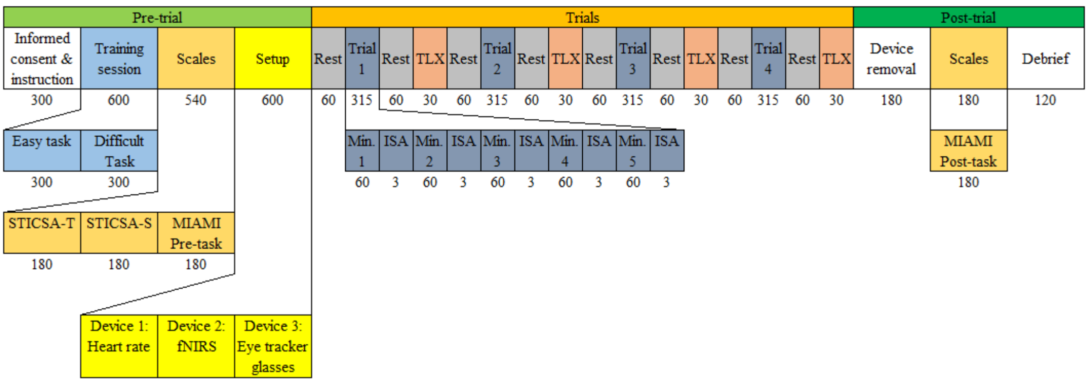


Figure 5.13: Procedure for Experiment 2.

Statistical analysis approach. The approach for performing statistical analysis was identical to Experiment 1 (see Section 5.3.1 p.113).

Synchronisation between Sensors and MATB. One of the challenges related to employing these sensors at the same time was time synchronisation, that is, starting and stopping the recording. From a practical side, it was unfortunately not possible to sync between these devices when an experiment session commenced. Regarding MATB, it cannot be integrated to external sensors due to coding limitation (Cegarra et al., 2020) thus it must be controlled manually by the researcher. Zephyr Bioharness 3, due to privacy concern, must be worn by participants in a private space. Consequently, the device must be started once it has been attached to participants' body. The researcher could not check visually whether the device had been 'turned on'. Regarding fNIRS and eye-tracking, they must be operated with their software running (Oxysoft and SMI Experiment Centre, respectively). To our best knowledge, there is no viable way to use these software packages under a single operating code. Therefore, all of these sensors and MATB were manually started and stopped during an experiment session. The solution for this particular challenge was to sync starting times of these sensors post-task. We used a code developed in Matlab (Mathworks Inc., Sherborn, MA, USA) to generate synchronised starting times of these sensors using inputs from respective recorded starting time. This strategy resulted in a single time stamp for each participant indicating the beginning of an experiment session, including for each treatment. From this time stamp, we could manually add the duration for each session. The time frame was used as the boundary for data that will be analysed.

5.4.2 Results

Testing hypothesis H5.2.1: subjective workload scores (NASA-TLX). NASA-TLX were used to capture participants' subjective perception towards the tasks. A 2 (low vs. high demand) x 7 (NASA-TLX dimensions and total scores) two-way repeated measures ANOVA with Bonferroni multiple testing correction method was performed to evaluate the effect of different dimensions of NASA-TLX tasks and demand levels on participants' subjective rating. There was a statistically significant interaction between demand levels and NASA-TLX dimensions on participants' subjective rating towards the tasks, F(2.96, 85.82) = 5.32, p = 0.002, $\eta_g^2 = 0.02$. Therefore, the effect of demand levels variable was analysed at each dimensions of NASA-TLX scales. Pairwise comparison shows that the mean subjective rating of workload was significantly different between low and high-demand task in all dimensions and total score, with effect size ranging from small to large (see Figure 5.14 and Table 5.13). Based on these results, we can support hypothesis H5.2.1.

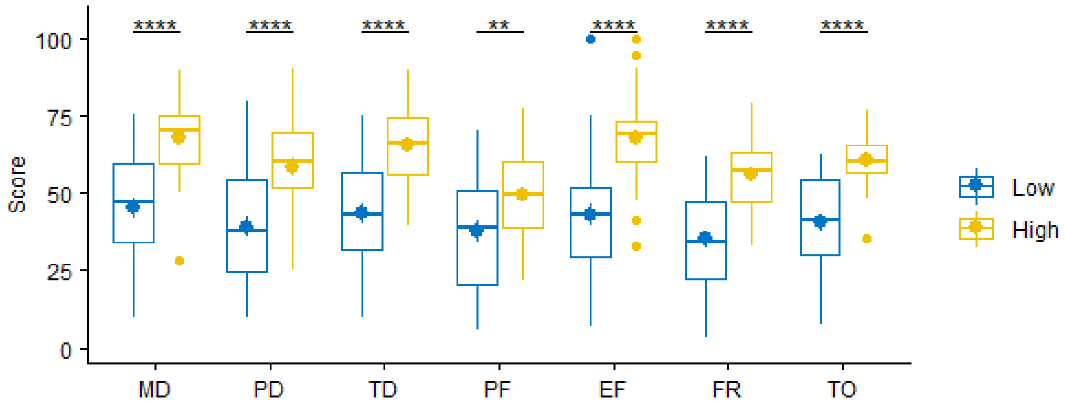


Figure 5.14: Pairwise comparisons of NASA-TLX (Experiment 2). The dots within the box indicate the mean from corresponding NASA-TLX scores, with standard error of the mean. For the total score of MWL, the difference is significant at p < 0.0001.

Table 5.13: Mean differences in NASA-TLX (Experiment 2).

Task	Level	Mean	SD	SE	p	Cohen's d
MD	Low High	45.183 67.633	17.947 12.933	3.277 2.361	p < 0.0001	d = 0.73
PD	Low High	39.067 58.400	17.087 15.992	3.120 2.920	p < 0.0001	d = 0.59
TD	Low High	43.350 65.700	17.818 12.235	3.253 2.234	p < 0.0001	d = 0.74
PF	Low High	37.867 49.400	19.716 14.064	3.600 2.568	p = 0.003	d = 0.34
EF	Low High	42.867 67.833	19.278 14.159	3.520 2.585	p < 0.0001	d = 0.75
FR	Low High	35.083 55.717	15.312 12.419	2.796 2.267	p < 0.0001	d = 0.75
TO	Low High	40.390 60.782	14.489 8.935	2.645 1.631	p < 0.0001	d = 0.86

Testing hypothesis H5.2.2: subjective workload scores (ISA). In this experiment, ISA was employed to cross-validate NASA-TLX. Pairwise t-tests revealed that there was significant difference in ISA score between low (M_{ISA} = 2.343, SD_{ISA} = 0.626) and high-demand tasks (M_{ISA} = 3.447, SD_{ISA} = 0.723), t(29) = -8.258, p < 0.001. The effect size for this difference was found to be considerable (Cohen's d = 1.633). These results support previous results in Experiment 1, that both demand levels subjectively differ, as captured by NASA-TLX and ISA. From these participants' subjective perception towards the tasks, we may conclude that task demand manipulation was successful. Figure 5.15 shows the results of ISA, and it suggests that hypothesis H5.2.2 can be supported.

Testing hypothesis H5.2.3: NASA-TLX-ISA correlation. A correlation analysis using the Pearson technique was performed to test the relationship between the two subjective MWL measurement. A large, positive, and significant correlation was found between NASA-TLX and ISA total scores ($r(60) = 0.644$, $p < 0.001$) thus hypothesis H5.2.3 can be supported. The result suggests that both subjective measurement methods agree to each other, supporting previous hypotheses that task demand manipulation worked. These facts may also suggest that the demand levels of the task practically generated different mental workload levels as perceived by participants.

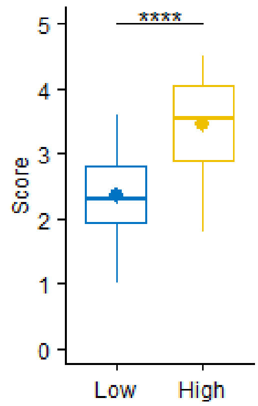


Figure 5.15: Pairwise comparisons of ISA (Experiment 2). The dots indicate the mean from ISA total score, with standard error of the mean. The difference is significant at $p < 0.001$.

Testing hypothesis H5.2.4: MATB performance. A performance score was created using the same method as in Experiment 1. A 2 (low vs. high demand) x 4 (MATB individual tasks and all tasks together) two-way repeated measures ANOVA with Bonferroni multiple testing correction method was performed to evaluate the effect of different MATB tasks and task demand levels on performance. There was statistically significant interaction between demand levels and MATB tasks on performance score, $F(3, 87) = 4.84$, $p = 0.004$, $\eta_g^2 = 0.02$. Therefore, similar to Experiment 1 results, the effect of demand levels variable was analysed at each MATB task. Pairwise comparison, using paired-sample t-test with Bonferroni correction, show that the mean performance score was significantly different between low and high-demand task in system monitoring task ($p < 0.001$), tracking task ($p < 0.001$), and all tasks together

(TOTAL) ($p < 0.001$), but not in resource management task (0.292). The effect size for these differences was found to be mostly small (see Figure 5.16 and Table 5.14).

Similar to Experiment 1, these results also suggest that both task demands practically differ, except in resource management task even though the difference in mean performance scores were similar to other tasks (higher in low demand task, vice versa). However, contrary to Experiment 1, the effect size may suggest that the difference between low and high-demand task appears to be small. From these performance scores, the conclusion is similar to Experiment 1, suggesting that a high-demand task results in poorer performance score because the task requires participants to deal with more task elements. Small difference in task performance score during low and high-demand tasks might be associated to proper training provided in this experiment. With proper training before the trial, participants could comprehend the task and subsequently set strategies for performing the task, particularly during high-demand level. Hypothesis H5.2.4, therefore, can be partially supported.

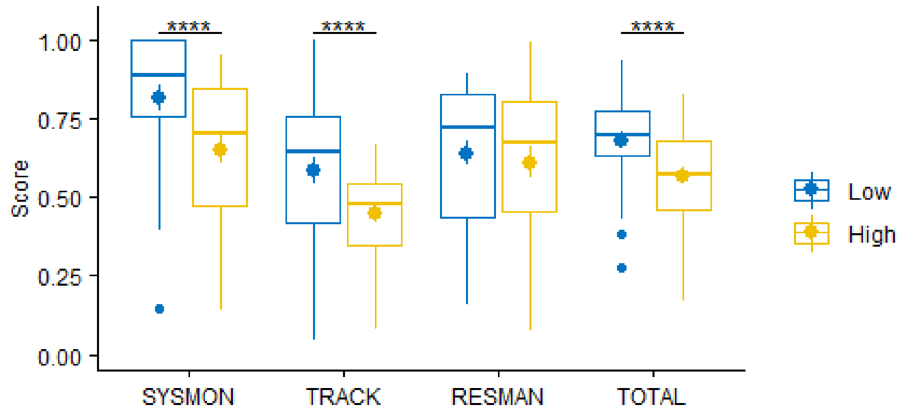


Figure 5.16: Pairwise comparisons of MATB performance scores (Experiment 2). The dots within the box indicate the mean from corresponding MATB task scores, with standard error of the mean. For the MATB total score, the difference is significant at $p < 0.0001$.

Table 5.14: Mean differences in MATB performance scores (Experiment 2).

Task	Level	Mean	SD	SE	p	Cohen's d
SYSMON	Easy Difficult	0.815 0.651	0.220 0.236	0.040 0.043	$p < 0.0001$	d = 0.37
TRACK	Easy Difficult	0.584 0.447	0.227 0.144	0.041 0.026	$p < 0.0001$	d = 0.37
RESMAN	Easy Difficult	0.639 0.609	0.216 0.272	0.039 0.050	$p = 0.292$	d = 0.06
TOTAL	Easy Difficult	0.680 0.569	0.150 0.155	0.027 0.028	$p < 0.0001$	d = 0.37

Testing hypothesis H5.2.5: brain activation. We followed identical preprocessing steps as in Experiment 1. Several participants data needed to be removed as they did not pass predefined correction and filtering criteria. More specifically, they were excluded as their raw signals exceeded channel pruning parameter (SNR threshold=2; dRange=1e-2 to 3e) (see Figure 5.17 for data exclusion flow and Table 5.15 for the excluded participants).

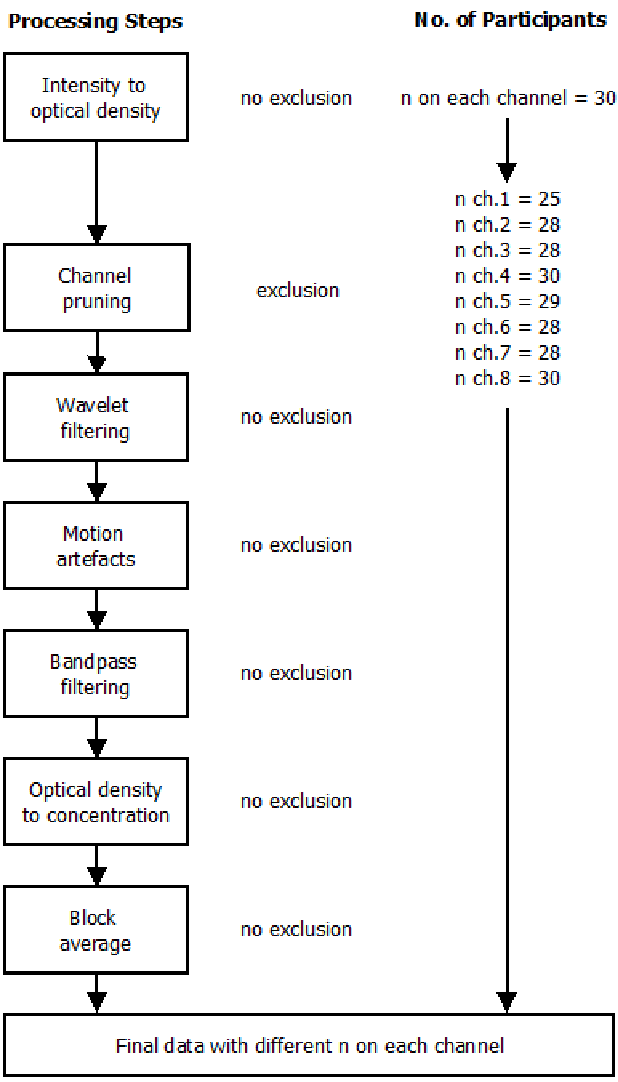


Figure 5.17: Data exclusion flow with the number of included participants for Experiment 2.

A one-sample t-test against zero to check the activation shows that all channels were activated during high-demand task, and Channel 1 and 2 were activated during low demand task. Table 5.16 shows the results of the t-tests for activated channels in Experiment 2.

Table 5.15: Participants whose data excluded in Experiment 2.

Channel	Participant's ID
Ch.1	3, 6, 16, 26, 30
Ch.2	16, 26
Ch.3	13, 26
Ch.4	n/a
Ch.5	26
Ch.6	7, 23
Ch.7	14, 30
Ch.8	n/a

Table 5.16: One-sample t-test results for activated channels (Experiment 2).

Level	Channel	Mean	SD	t statistics	p
Low	Ch.1	0.129	0.211	t(24) = 3.053	0.006
	Ch.2	0.122	0.288	t(27) = 2.239	0.034
High	Ch.1	0.149	0.245	t(24) = 3.029	0.006
	Ch.2	0.278	0.333	t(27) = 4.413	0.000
	Ch.3	0.214	0.287	t(27) = 3.955	0.000
	Ch.4	0.148	0.279	t(29) = 2.899	0.007
	Ch.5	0.139	0.227	t(28) = 3.307	0.002
	Ch.6	0.151	0.283	t(27) = 2.811	0.009
	Ch.7	0.228	0.294	t(27) = 4.111	0.000
	Ch.8	0.224	0.325	t(29) = 3.768	0.000

A 2 (low vs. high demand) x 8 (Channel 1 to 8) repeated measures ANOVA with Bonferroni correction was used to compare brain activation (HbDiff) of participants during two task demand levels of MATB tasks. We found significant interaction between demand levels and fNIRS channels on brain activation, F(7, 203) = 2.4, p = 0.022, $\eta_g^2 = 0.007$. Since according to the calculation it violates the sphericity assumption, Greenhouse-Geisser correction was applied. However, we did not find any significant interactions (F(4.31, 86.29) = 2.09, p = 0.085, $\eta_g^2 = 0.007$).

Pairwise comparison, using paired-sample t-tests, revealed the mean brain activation was significantly different in Channel 6 ($p < 0.035$) and Channel 7 ($p < 0.011$). The results suggest that in Channel 6 high-demand task tends to activate corresponding areas of brain ($M_{act.}$ = 0.151, $SD_{act.}$ = 0.283) than low demand task ($M_{act.}$ = 0.015, $SD_{act.}$ = 0.326). Moreover, similar difference in activation was also found in Channel 7 (high-demand task: $M_{act.}$ = 0.228, $SD_{act.}$ = 0.294; low demand task: $M_{act.}$ = 0.043, $SD_{act.}$ = 0.326). The effect size in these two channels were reported to be small. Based on the results, we may partially support hypothesis H5.2.5. Figure 5.18 and Table 5.17 shows the results from the hypothesis testing of brain activation changes on each channel.

Table 5.17: Mean differences in brain activation scores (Experiment 2).

Channel	Level	Mean	SD	SE	p	Cohen's d
Ch.1	Low	0.107	0.198	0.036	p = 0.783	d = 0.04
	High	0.124	0.230	0.042		

(continued on next page)

Table 5.17, continued

Channel	Level	Mean	SD	SE	p	Cohen's d
Ch.2	Low High	0.114 0.259	0.280 0.329	0.051 0.060	p = 0.076	d = 0.26
Ch.3	Low High	0.084 0.200	0.326 0.282	0.060 0.051	p = 0.136	d = 0.20
Ch.4	Low High	0.092 0.148	0.328 0.279	0.060 0.051	p = 0.385	d = 0.01
Ch.5	Low High	0.057 0.135	0.298 0.225	0.054 0.041	p = 0.223	d = 0.15
Ch.6	Low High	0.014 0.141	0.315 0.276	0.057 0.050	p = 0.035	d = 0.23
Ch.7	Low High	0.040 0.213	0.315 0.289	0.057 0.053	p = 0.011	d = 0.31
Ch.8	Low High	0.088 0.224	0.314 0.325	0.057 0.059	p = 0.065	d = 0.22

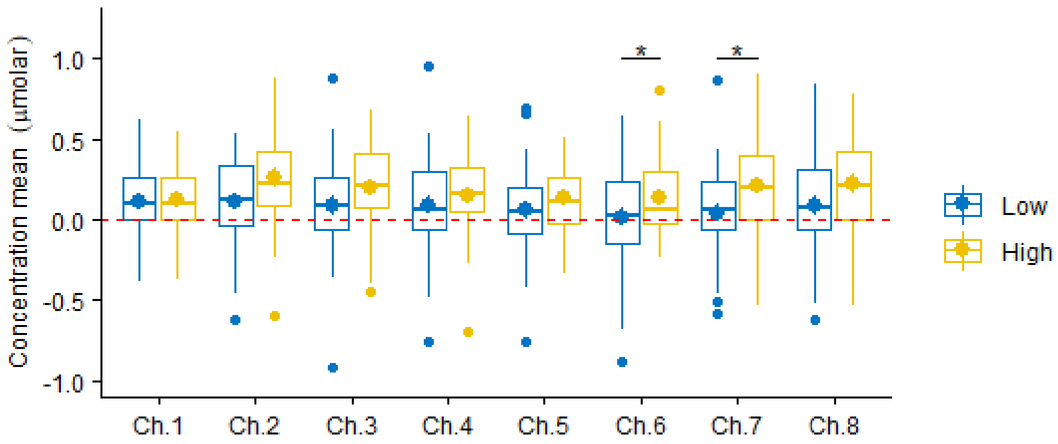


Figure 5.18: Comparisons of brain activation between channels (Experiment 2). The dots within the box indicate the mean from brain activation index with standard error of the mean. The more positive values indicate the more brain activation and vice versa. The differences in both channel 6 and 7 are significant at $p < 0.05$.

Testing hypothesis H5.2.6: heart rate measures. Heart rate analyses were performed using Kubios HRV Standard Version 3.3.1 for Windows 64-bit from Kubios company (Kuopio, Finland). We decided to apply radical steps of artefact correction, meaning that a 'very strong' threshold logarithm was applied in first step before the weaker one. This is because, for privacy concerns, participants had to wear the device themselves in a private space such as a restroom. The instruction from the device user's manual as well as detail explanation by researcher were also provided. However, we need to anticipate that some participants might not correctly wear the device, and it will affect the recording. Therefore, a conservative approach seemed to be the most practical way to retain as many clean signals as possible. The correction was made possible by comparing every inter-beat interval (IBI) values against a local average interval, with different threshold ranging from very low (0.45 seconds) to very strong (0.05 seconds). If the amount of corrected signals were high, or above 5% from the sample signals, we applied the less powerful threshold until we obtained sufficient signals (artefacts < 5% of the sample). Using this strategy, data from seven participants had to be removed due to a considerable amount of noise.

We first checked differences between heart rate measures (rate, RMSSD, SDNN) and task demand levels using pairwise t-test. Raw heart rate (measured in beat per minute) of participants were found higher when performing high-demand task (M_{hr} = 81.36, SD_{hr} = 10.70) than low demand task (M_{hr} = 80.15, SD_{hr} = 11.27). Whereas, HRV measures (RMSSD and SDNN, both measured in ms) were acting oppositely. Participants' HRV was lower in high-demand task (M_{rmssd} = 29.84, SD_{rmssd} = 16.72; M_{sdnn} = 30.46, SD_{sdnn} = 13.49) than low demand task (M_{rmssd} = 53.46, SD_{rmssd} = 10.74; M_{sdnn} = 30.59, SD_{sdnn} = 13.17). However, we found significant differences in heart rate ($t(22) = -2.531$, $p = 0.019$) and RMSSD ($t(22) = 8.113$, $p < 0.001$), but not in SDNN ($t(22) = 0.122$, $p = 0.904$), see Figure 5.19. Despite a significant difference in heart rate measure, the effect size is trivial (Cohen's d = 0.06). Meanwhile, the effect size for difference in RMSSD was found to be large (Cohen's d = 0.86). Based on the results, we may partially support hypothesis H5.2.6.

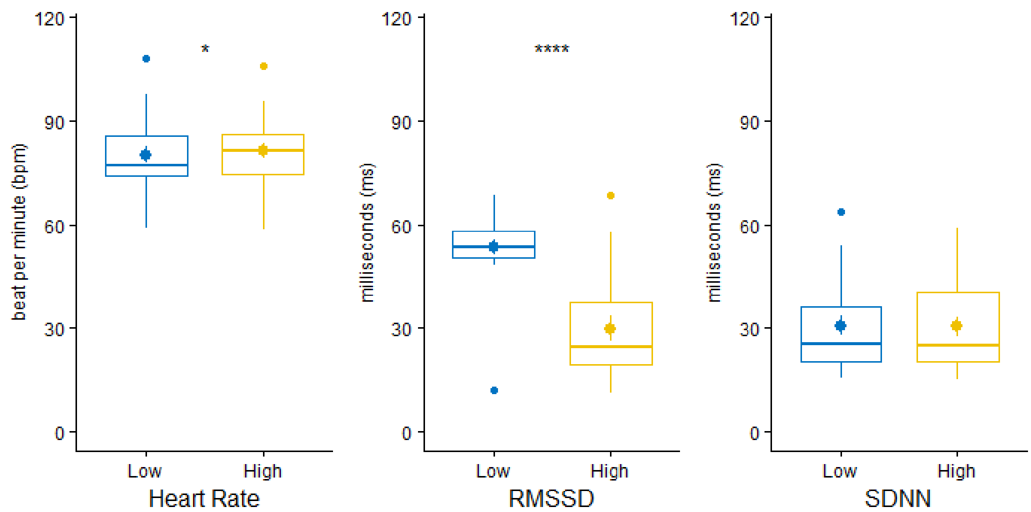


Figure 5.19: HR measures differences indicating mental workload changes. The dots within the box indicate the mean from corresponding HR measures, with standard error of the mean. Heart rate is expressed in beat per minute (bpm) while RMSSD and SDNN are expressed in milliseconds (ms). The significant differences are found in HR at $p < 0.05$ and RMSSD at $p < 0.001$.

Testing hypothesis H5.2.7: pupil dilation. Data from the eye-tracker device was live saved to the computer during the experiment and can be directly exported into

CSV format afterwards. During recording, data from seven participants were found to be corrupted, and thus they had to be excluded from analysis. The remaining data were then manually marked for the starting point based on the generated synchronised time stamp (see Section 5.4.1 regarding the synchronisation strategy). We then examined differences between participants' pupil size in both eyes and task demand levels using pairwise t-tests. Pupil size in both eyes dilated wider when performing high-demand task (left eye: M_{left} = 4.008, SD_{left} = 0.752; right eye: M_{right} = 4.268, SD_{right} = 0.719) than low demand task (left eye: M_{left} = 3.852, SD_{left} = 0.756; right eye: M_{right} = 4.042, SD_{right} = 0.700). The results also showed significant differences in both eyes (left eye: $t(22) = -2.941$, $p = 0.008$; right eye: $t(22) = -5.214$, $p < 0.001$). Despite significant differences, the effect size for both eyes were found to be small (left eye: Cohen's d = 0.11; right eye: Cohen's d = 0.16). Figure 5.20 shows the results from the hypothesis testing of pupil dilation on both eyes. Based on these results, we can support hypothesis H5.2.7.

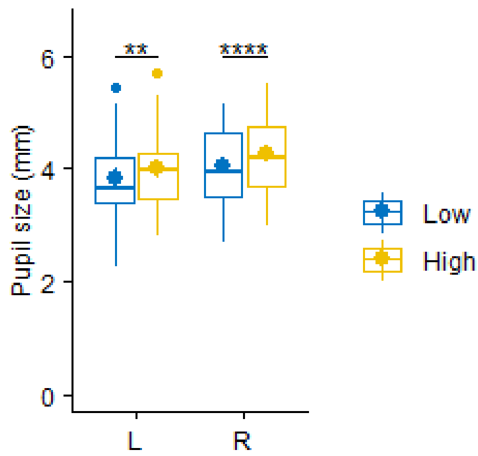


Figure 5.20: Pupil dilation differences indicating mental workload changes. The dots indicate the mean of pupil diameter (measured in millimetres/mm) with standard error of the mean. The differences are significant in both eyes, at p < 0.01 for the left eye and p < 0.001 for the right eye.

Testing hypothesis H5.2.8-H5.2.11: correlations between TLX with performance and physiological measurements. For each hypothesis tested, a partial correlation was run to determine the relationship between subjective MWL total scores and performance (H5.2.8), brain activation (H5.2.9), HR/HRV (H5.2.10), and pupil dilation (H5.2.11), whilst controlling for anxiety and motivation. Figure 5.21 compiles the results of correlations between variables in Experiment 2 (including H5.2.3 result).

Regarding performance, a small, negative, and significant correlation was found between subjective MWL and MATB total scores ($r(60) = -0.274$, $p = 0.039$). Meanwhile, correlations between subjective MWL and the elements of MATB were found to be small, negative, and not significant, i.e. SYSMON (r(60) = -0.196, p = 0.144); TRACK (r(60) = -0.201, p = 0.135); RESMAN (r(60) = -0.200, p = 0.136). These results suggest that, in this experiment, perceived workload of the task can be confirmed by performance total scores. However, while the relationship direction was consistent with the hypothesis, the individual element of the MATB did not show significant correlation with subjective MWL scores. Based on these results, we may partially support hypothesis H5.2.8.

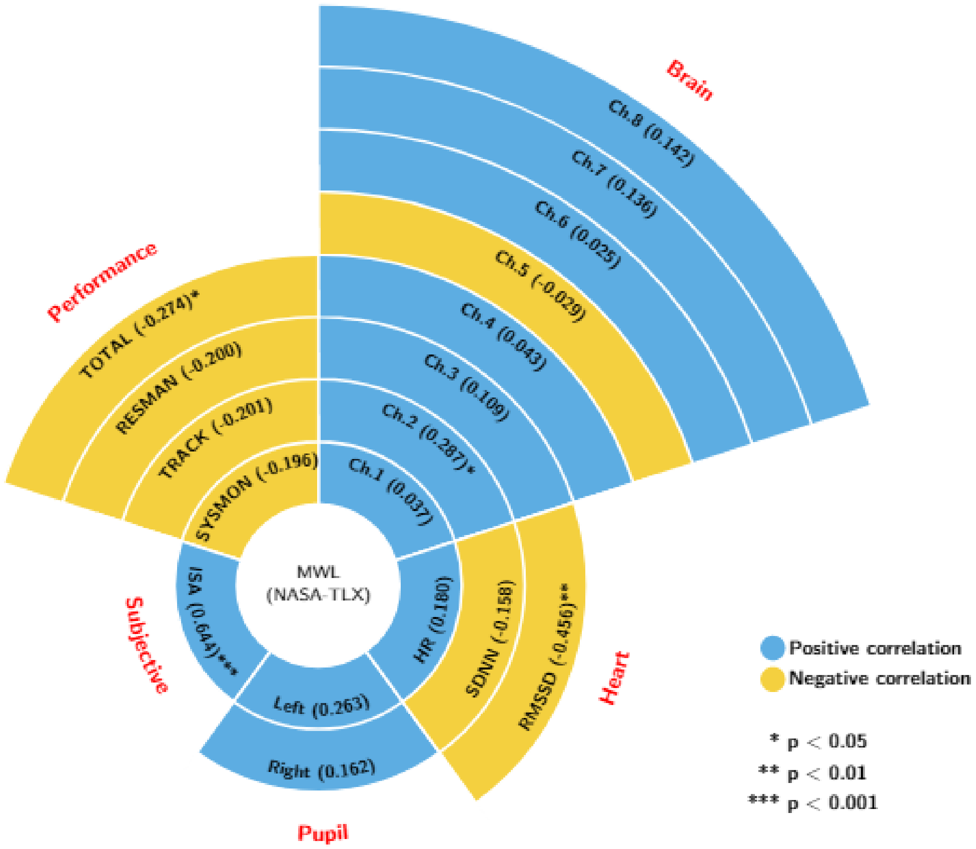


Figure 5.21: First-order correlation between variables in Experiment 2. The blue shading indicates positive correlation, while the amber shading indicates negative correlation. The number inside the bracket indicates the magnitude of the correlations between subjective MWL score/NASA-TLX and corresponding variables, ranging from -1 to 1. Significant results are indicated by asterisk annotation.

Concerning brain activation, a small, positive, and significant correlation was found between subjective MWL score and Channel 2 only (r(60) = 0.287, p = 0.03). However, correlating subjective MWL scores with the rest of the channels showed a small (Channel 3, 7, 8) or even weak (Channel 1, 4, 5, 6) correlations. Except for Channel 5, the direction of the relationships was positive as hypothesised. The results may suggest that brain activation change might be able to indicate MWL changes, but only in Channel 2. Based on the results, hypothesis H5.2.9 was therefore partially supported.

Regarding heart rate measurements, the direction of the relationships varied. Heart rate had a small, positive, but not significant correlation with subjective MWL score ($r(46) = 0.180$, $p = 0.249$); whilst SDNN and RMSSD had a negative correlation with subjective MWL score. Furthermore, SDNN had a small correlation ($r(46) = -0.158$, $p = 0.311$) yet RMSSD had a medium and significant correlation ($r(46) = -0.456$, $p = 0.002$). While the directions of the relationship were as hypothesised, significant correlation was only found in RMSSD. The results indicate that perceived workload change can be confirmed by changes in RMSSD. Thus, hypothesis H5.2.10 can be partially supported.

The last correlation analysis was between subjective MWL score and pupil dilation in both eyes. A small, positive, but not significant correlation was found between these two variables, i.e. with right eye ($r(46) = 0.162$, $p = 0.299$) and with left eye ($r(46) = 0.263$, $p = 0.089$). The direction of the relationship might suggest that perceived MWL can be confirmed by changes in pupil diameter. However, since the correlations were not significant, this conclusion has to be limited. Hypothesis H5.2.11 cannot therefore be supported. Regarding the role of psychological variables, we did not find much different between zero- and first-order correlation in all variables of interest in terms of their direction, magnitude, and significance levels. Thus, it can be said that in Experiment 2 anxiety and motivation had very little influence in controlling the relationship between variables of interest.

5.4.3 Discussion

This experiment aimed to verify the results obtained from Experiment 1 by refining the methodology and techniques of the measurements. It also aimed to broaden the scope of the MWL measurements by employing other physiological devices. In Experiment 2, correlations between these physiological measurements and subjective MWL measure were also performed to confirm whether they can inform changes in MWL during MATB task. The role of psychological conditions during the task was also considered in this experiment. The summary of hypotheses testing results are presented in Table 5.18.

Since Experiment 2 was theoretically a replication of Experiment 1, the results from Experiment 2 will be discussed in the manner of comparison with the results from the previous experiment. Table 5.19 presents the comparison of the results between Experiment 1 and Experiment 2.

Table 5.18: Hypotheses testing results for Experiment 2.

No.	Hypothesis (alternative) statement	Hypothesis testing
H5.2.1	NASA-TLX scores in all dimensions will score higher in high-demand task than low demand task.	Supported
H5.2.2	ISA scores in all dimensions will score higher in high-demand task than low demand task.	Supported
H5.2.3	NASA-TLX will positively correlate with ISA.	Supported
H5.2.4	MATB scores in all tasks will score lower in high-demand task than low demand task.	Partially supported
H5.2.5	Brain activation in all fNIRS channels will score higher in high-demand task than low demand task.	Partially supported
H5.2.6	Heart rate will score higher in high-demand task than low demand task, whilst its variability measures will score oppositely.	Partially supported
H5.2.7	Pupil diameter in both eyes will dilate wider in high-demand task than low demand task.	Supported
H5.2.8	Controlling state and trait anxiety and motivation, MATB score will negatively correlate with NASA-TLX score.	Partially supported
H5.2.9	Controlling state and trait anxiety and motivation, brain activation will positively correlate with NASA-TLX score.	Partially supported
H5.2.10	Controlling state and trait anxiety and motivation, HR will positively correlate with NASA-TLX score; whilst RMSSD and SDNN will negatively correlate with NASA-TLX score.	Partially supported
H5.2.11	Controlling state and trait anxiety and motivation, pupil diameter will positively correlate with NASA-TLX score.	Not supported

Regarding the subjective MWL measure and MATB performance, there were not many differences between these two experiments. In Experiment 2, all dimensions of NASA-TLX were found to be significantly different, suggesting that the task was practically generating different mental workload. Thus, participants perceived them, in terms of their workload level, differently. From these results as well, we may conclude that task manipulation was successful. Physical demand dimension, however, was perceived differently between demand levels in Experiment 2, contrary to Experiment 1 that tends to be not much different. This discrepancy might be interpreted as the result of task design change. While the demands of the task did not change, participants experienced the actual task with both demand levels during the training session. In total, participants played six five-minute trials of the MATB task, with two of them for training purposes. This was different to Experiment 1 that had only two-minute training for each demand levels. We therefore presumed that physical effort during this experiment session were higher than the previous one, and thus affected the way participants perceived physical demands of the task.

Table 5.19: The results' comparison of Experiment 1 and 2.

Variables/factors	Experiment 1			Experiment 2		
	Significance	Cohen's d	Effect Size	Significance	Cohen's d	Effect Size
MATB/system monitoring	****	0.35	small	****	0.37	small
MATB/tracking	****	0.83	large	****	0.37	small
MATB/resource manage.	n/s			n/s		
MATB/total performance	****	0.57	medium	****	0.37	small
TLX/mental demand	**	0.28	small	****	0.73	medium
TLX/physical demand	n/s			****	0.60	medium
TLX/temporal demand	***	0.39	small	****	0.74	medium
TLX/performance	*	0.20	small	**	0.34	small
TLX/effort	***	0.33	small	****	0.75	medium
TLX/frustration	**	0.24	small	****	0.75	medium
TLX/total workload	***	0.37	small	****	0.86	large
fNIRS/channel 1	n/s			n/s		
fNIRS/channel 2	n/s			n/s		
fNIRS/channel 3	n/s			n/s		
fNIRS/channel 4	n/s			n/s		
fNIRS/channel 5	n/s			n/s		
fNIRS/channel 6	n/s			*	0.21	small
fNIRS/channel 7	n/s			*	0.29	small
fNIRS/channel 8	n/s			n/s		
Heart/beat	n/a	n/a	n/a	*	0.05	trivial
Heart/RMSSD	n/a	n/a	n/a	****	0.86	large
Heart/SDNN	n/a	n/a	n/a	n/s		
Eye/left	n/a	n/a	n/a	**	0.11	trivial
Eye/right	n/a	n/a	n/a	****	0.16	trivial
ISA	n/a	n/a	n/a	****	1.63	very large

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$; n/a = not applicable; n/s = not significant

In this experiment, ISA was also employed to cross-validate NASA-TLX. Specifically, ISA can show whether MWL was subjectively experienced differently throughout the task according to their demand levels. It was considered essential because NASA-TLX was administered after the task, thus prone to misjudgement. The results yielded consistent scores with NASA-TLX, suggesting significant difference in demand levels from both tasks were valid. A positive, high, and significant correlation between these two subjective measurements (hypothesis H5.2.3) supports this conclusion. Furthermore, the effect size for ISA was also found to be very large, while the effect for NASA-TLX (total unweighted score) was large. These results suggest that the demands induced to the MATB tasks were evidently different. In other words, participants would perceive the high-demand task as a more 'difficult' task. The effect size of NASA-TLX scores in this experiment was higher than in Experiment 1. This may suggest that participants understood the task and this may be attributed to the proper pre-task training provision. In Experiment 1, we suspected that participants responded to the stimuli during the task arbitrarily due to lack of training.

MATB performance score for this experiment was identical to Experiment 1. Resource management task score showed no difference as in the previous experiment, suggesting the task did not tax much attentional resource as in the other two subtasks. This, as explained in the Experiment 1 discussion, might be due to the nature of the task that tends to be not-so-dynamic. Participants might be able to leave their attention away from this task once they believe they have applied a 'correct' strategy in balancing the fuel tanks. In this Experiment, MATB performance score can confirm the mental workload level generated by the task. The effect size for MATB total performance scores in Experiment 2 was small, while in Experiment 1 was medium. We interpreted that in Experiment 2, as the results of better task comprehension, participants might have handled the high-demand task more strategically. This consequently yielded performance scores that were not too different between low and high-demand task. Contrary to this, in Experiment 1, the performance score differences were more evident. This may be attributed to the inability to handle the high-demand task more strategically, resulting in considerable lower scores. Combining MATB performance score and subjective MWL scales (NASA-TLX and ISA), it can be confidently said that the demands within the task had generated practically different levels of mental demands. The difference, furthermore, may serve as the essential basis for physiological measurements of MWL. In other words, the tasks have been proven to be different in terms of demand levels, and now it is the matter of whether the physiological devices can detect them.

Concerning brain activation, the task demand changes did apparently have an effect on fNIRS device. In this experiment, fNIRS can detect brain activation changes during low and high-demand tasks. This notion was supported by the results of the t-tests against zero, suggesting activation in all channels when performing a high-demand task; and in Channel 1 and 2 when performing a low demand task. We might say, based on this result, that the high-demand tasks in particular have successfully triggered certain cognitive function and the brain responded it by requesting more blood containing oxygen for producing energy within the neurons. Except for Channel 1, however, fNIRS has actually shown changes in concentration, yet the magnitude of these changes seem to be insufficient to be considered 'significantly different' from zero. This might subsequently affect the analysis of the concentration differences between low and high-demand tasks. From the results, significant difference between low and high-demand levels were found only in Channel 6 and 7, corresponding to Brodmann Area 10 and 46 of the left part of PFC. The effect size was found to be small, indicating that the differences was not too evident. We may suggest that the oxygenation in the PFC region occurred almost similarly when responding to the low and high-demand tasks. Similar to activation in a low demand task, from Figure 5.18, we might have a sense of seeing differences in concentration changes between low and high-demand tasks in almost all channels. Nevertheless, they were statistically insufficient to infer true differences. We, therefore, might say that fNIRS was not sufficiently sensitive to detect such a task battery. It might also be said that fNIRS was only suitable to detect task with larger difference in demand levels (Argyle et al., 2021). Consequently, we cannot confidently say that fNIRS is a robust way to detect MWL changes in task battery such as MATB.

The remaining variables to be discussed were peculiar to this experiment, thus not comparable to the previous experiment. Two physiological measuring devices were applied in Experiment 2 (heart rate and pupil dilation sensors). We were interested in the basic measures of heart rate activities, such as raw rate and several variability measures. Heart rate and one of the variability measure of HR (RMSSD) showed significantly different score between low and high-demand task. However, the effect size for the heart rate was trivial. This means that the differences, despite its significant results, was barely distinguishable. This may be attributed to the range of the heart beat that tends to be narrow in most healthy people when performing cognitive tasks. In other words, completing a high-demand task would not cause an extreme spike of the beat. Nevertheless, when it comes to HR variability (RMSSD), the differences became evident, as indicated by a large effect size. Meanwhile, SDNN also showed difference yet not significant statistically. In normal conditions, HR and RMSSD acted oppositely and, in the context of MWL measurement, a task with higher MWL would theoretically trigger higher heart rate but lower HR variability scores. We observed increase in participants' heart rate along with decrease in their RMSSD score during a MATB task with higher demand.

Similar to HR, pupil dilation measurement in both eyes also showed significant differences between low and high-demand tasks. A task with higher MWL would trigger pupil dilation. In this experiment, as hypothesised, participants' pupils in both eyes were observed to be wider when performing a high demand MATB task. However, the effect size was trivial. Similar to HR, this might be attributed to the nature of pupillary movement that was slight and limited. In other words, in most healthy people, the pupils cannot dilate any wider than their normal range. Therefore, the difference in pupil dilation in response to low and high-demand task would be barely distinguishable.

These phenomena were made possible to be observed due to the connection of heart and pupillary muscles to the autonomic nervous system. It is argued that there is a pathway from the brain to regulate heart activity through both sympathetic and parasympathetic nerves (Thayer & Lane, 2009). Beside heart, pupillary muscles are said to be connected to the brain regions that are responsible for cognitive and autonomic functions via both sympathetic and parasympathetic nerves activation (Eckstein et al., 2017). Based on these results, we may partially support hypotheses related to these additional physical measurements. Hence, we may conclude that MWL changes during MATB tasks can be confirmed by changes in HR and RMSSD as well as pupil dilation.

Regarding correlations, as in Experiment 1, we put subjective measurements of MWL as captured by NASA-TLX scale as the centre of interest. We assume the notion that MWL is argued to have a strong element of subjectivity (Van Acker et al., 2018). MWL is an intangible phenomenon, and thus the way people report their experience regarding the task seems to matter the most. Therefore, confirming physiological measures to the subjective one could provide further confidence to the application of such measurement methods. From Figure 5.21 the direction of the correlations between these variables and NASA-TLX score were as hypothesised. The magnitude of the correlations ranges from low to medium, with the exceptions in several fNIRS channels. However, significant correlation were only found in total scores of MATB performance, fNIRS Channel 2, and RMSSD.

Looking at these results, the correlations might suggest that physiological measurements applied in this experiment converge towards the subjective MWL experience. Combined with aforementioned data of these physiological measurements in detecting MWL changes, these correlational results thus can be carefully interpreted as a successful confirmation of the methods. Regarding the correlation, the issue might be more related to statistics. Correlation coefficient is said to be arbitrary and influenced by many statistical properties such as variability, sample size, and the presence of outliers (Goodwin & Leech, 2006). In our case, we did find outliers, particularly in fNIRS data. The variability of the data, in fNIRS for example, were also observed to be high due to the nature of the measurement that was sensitive to artefacts such as body movement. The sample size in Experiment 2 might also be considered insufficient. More specifically, approximately a third of the data from heart rate and pupil dilation had to be excluded due to recording errors.

Partial correlation analysis in this experiment suggested that there was no difference between zero- and first-order correlation, meaning that control variables did not much influence the results. The results suggested that during MATB performance, changes in MWL as captured by physiological measures and performance score were not dependent on the anxiety and motivation of participants. The explanation for this result can be twofold. Firstly, the scale seems to be designed for a clinical situation in which people might have consistently felt anxious and unmotivated. Secondly, we did not manipulate participants' anxiety and motivation in both experiments. Thus, the score obtained from both scales came from common individuals who were most likely to be not extremely anxious or unmotivated. This might then affect the ability to explain the roles of these two variables in controlling the relationships in both Experiment 1 and 2. Despite their effects on changing the correlation magnitude in system monitoring and MATB total score in Experiment 1, the significance levels did not change. It may be therefore safe to say that, in our case, MWL changes were not affected by anxiety and motivation level.

Concluding Experiment 2, MWL changes during a MATB task might be able to be detected using fNIRS, pupil dilation sensor, and heart rate monitoring device. These physiological measures tend to converge to the subjective experience of MWL as reported by participants. However, since the statistical significances were not found uniformly, further investigation is necessary to infer more confidently.

5.5 Chapter Summary

This chapter consisted of two experiments aiming to investigate whether mental workload during a simulated flying task can be detected by physiological indicators. At first, we examined the ability of fNIRS and found that the device was not sensitive enough for such a task. Issues regarding the task design were also discussed and led to the necessity to undertake another experiment with several changes in the design of the task. Experiment 2 added two other physiological measurements to detect changes in MWL and a subjective MWL scale to cross-validate the existing NASA-TLX. The results suggest that MWL changes can be detected by partial channels of fNIRS, pupil dilation sensor in both eyes, and heart rate with one of its variability measures, in this case, RMSSD. These measurements tend to converge to the subjective MWL measurement yet interpretation regarding this matter must be done carefully due to inconsistent statistical significance levels found on some relationships. Anxiety and motivation of participants were apparently not influential to MWL changes. The next chapter will investigate whether MWL can be predicted pre-task.

Chapter 6

Study 4: A Preliminary Investigation on Vicarious Observation of Mental Workload during a Simulated Flying Task

6.1 Chapter Overview

This chapter presents an experiment study investigating if MWL when performing a simulated flying task can be predicted before the actual task. This chapter addresses the fifth research aim, which is 'to investigate whether MWL changes during a flying task can be predicted before the task'. An online experiment was applied to gather participants' responses to subjective MWL of MATB tasks by merely carefully observing the videos of the task. In this study, participants did not directly experience the task. This is a preliminary study to understand expectations about task demands in the context of flying task. This study therefore might provide additional explanation of MWL measurement during a simulated flying task such as MATB. This chapter also concludes the empirical research in this thesis.

6.2 Introduction

From the previous three studies, we have attempted to understand MWL in the context of flying tasks from different perspectives. The most common way of studying

MWL tends to be homogeneous. Participants are asked to experience certain task(s) and then asked whether the task has kept them 'busy' enough. Theoretically speaking, this is the way to measure how many cognitive resources that have occupied and left when performing the task(s). MWL studies also come with a more objective way to measure cognitive resources' occupancy. As we have attempted in Study 3, various physiological measuring devices could provide insightful information about mental workload. However, there is a less-explored area that appears to be useful in leveraging our understanding of MWL, particularly in the context of flying task. This, more specifically, leads to a research question i.e. *"can we predict MWL of a simulated flying task prospectively?"*. In other words, if we are given a description of a task in a visual form, for example, can the demands thus MWL of the task be distinguished accurately?

Few studies have been undertaken to explore this area. The studies from Sublette et al. (2009, 2010), for example, attempted to see the difference between prospective and retrospective evaluations of MWL in performing medical surgery task. The results from their experiments suggested that expectation about task demand tends to vary among participants. This was said to be dependent on the dimensions of MWL being considered. In their case, furthermore, prospective judgement of MWL can serve as a reliable proxy to retrospective judgment when the key components of the overall workload are physical and temporal demands. This kind of study can provide insights about the way participants see the task and its components. Subsequently, it can lead to determine more proper strategies or approaches to complete the real task.

We humbly say that this would be a preliminary attempt to explore this area, particularly in the context of flying task. The results from this study might add valuable explanation in understanding MWL during a simulated flying tasks. With the previously mentioned research question, this study therefore aims to investigate if MWL can be predicted pre-task. A simple online experiment was undertaken to achieve the study goal. Our hypothesis for this study was:

1. When presented with merely the videos of the task, participants will rate the subjective workload of the task with high demand higher than the task with low demand (H6.1.1).

6.3 Methods

Experiment task and design. The design of this study was repeated measures. The independent variable was task demand levels of MATB task: low and high. The tasks were separately presented, starting from system monitoring (SYSMON), tracking (TRACK), resource management (RESMAN), and all tasks together (MULTI). In terms of the task demand, each level was presented once with easy to difficult order. The task lasted for one minute each. The dependent variables were subjective workload ratings, measured by NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988).

Participants. Twenty-one participants were recruited from the university students and staff (M_{age} = 33.43, SD_{age} = 6.56) by using various means such as institutional emails and groups (Teams) with the link of the study included. However, personal approach, i.e. by directly contacting colleagues or fellows were also carried out to maximise the availability of participants. Ethics approval for this study was granted from Faculty of Engineering Research Ethics Committee, the University of Nottingham.

Apparatus. This study was conducted online using Microsoft Forms consisting of sample videos of MATB tasks and NASA-TLX scales.

Procedure. During their study appointment, participants were able to complete the study online, including the informed consent form and several demographic questions. The experiment was started by the appearance of an easy MATB task video with a particular order: SYSMON, TRACK, RESMAN, and MULTI. After watching each video that lasted for one minute, participants were asked to complete NASA-TLX scales before continuing to the next video. The low demand tasks sequence was then followed by sequence of high demand MATB task with identical order and protocol. A monetary compensation was then provided as reward for their participation on the study.

6.4 Results

System monitoring task. A paired-samples t-test with Bonferroni correction was conducted to compare scores from all dimensions of NASA-TLX scales in low and high-demand level of system monitoring task. There was significant difference in mental demands (MD), physical demands (PD), frustration (FR), and total scores (TO) of NASA-TLX scales. We, however, did not find significant difference in temporal demands (TD), performance (PF), and effort (EF) (see Table 6.1for the statistics and Figure 6.1 for the graphs).

Table 6.1: The results of t-tests for system monitoring task.

Dimension	Level	Mean	SD	t(20) = ?	p
MD	Low High	5.286 6.333	2.610 2.176	-2.137	0.045*
PD	Low High	4.095 5.571	2.587 2.378	-3.239	0.004**
TD	Low High	5.524 6.476	2.822 1.990	-1.493	0.151
PF	Low High	3.333 4.381	1.770 1.717	-1.891	0.073
EF	Low High	6.190 6.333	2.804 2.058	-0.292	0.773
FR	Low High	4.857 5.810	2.833 2.482	-2.118	0.047*
TO	Low High	4.881 5.817	2.134 1.733	-2.259	0.035*

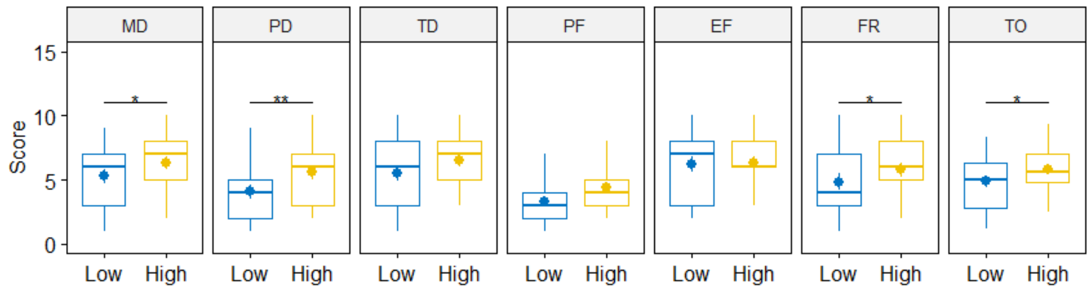


Figure 6.1: System monitoring task scores between low and high-demand level. The dots within the box indicate the mean of the scores, with standard error of the mean. For MD, FR, and total score of the scale, the differences are significant at $p < 0.05$ while for PD at $p < 0.01$.

These results suggest that task demand levels of system monitoring task can be partially differentiated accurately before the task. Specifically, our results suggest that participants can subjectively perceive the difference between easy and difficult task in terms of its mental demand, physical demand, frustration level, and total workload score.

Tracking task. A paired-samples t-test with Bonferroni correction was also conducted to compare scores from all dimensions of NASA-TLX scales in low and high-demand level of tracking task. However, significant difference of NASA-TLX scores when seeing low and high-demand task was only found in frustration level (FR) (see Table 6.2 for the statistics and Figure 6.2 for the graphs).

Table 6.2: The results of t-tests for tracking task.

Dimension	Level	Mean	SD	t(20) = ?	p
MD	Low High	4.143 4.857	2.081 2.287	-1.826	0.082
PD	Low High	3.905 4.238	2.234 2.508	-1.022	0.319
TD	Low High	4.048 4.524	2.224 2.421	-1.140	0.268
PF	Low High	2.429 2.667	1.777 1.065	-0.527	0.604
EF	Low High	4.429 4.524	2.315 2.316	-0.257	0.800
FR	Low High	3.381 4.190	1.746 2.442	-2.968	0.008**
TO	Low High	3.722 4.167	1.654 1.978	-1.684	0.108

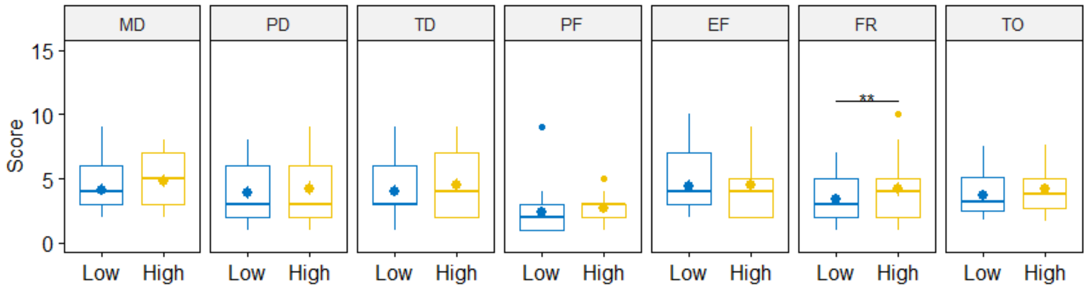


Figure 6.2: Tracking task scores between low and high-demand level. The dots within the box indicate the mean of the scores, with standard error of the mean. For FR score of the scale, the difference is significant at $p < 0.01$.

As in the system monitoring task, the results suggest that demand levels of the tracking task can be partially differentiated accurately before the task. Specifically, our results suggest that participants can subjectively perceive the difference between low and high-demand task only in frustration level.

Resource management task. A paired-samples t-test with Bonferroni correction was also conducted to compare scores from all dimensions of NASA-TLX scales in low and high-demand level of resource management task. In this task, there was no significant difference of NASA-TLX scores when seeing low and high-demand tasks (see Table 6.3 for the statistics and Figure 6.3 for the graphs).

Table 6.3: The results of t-tests for resource management task.

Dimension	Level	Mean	SD	t(20) = ?	p
MD	Low High	5.952 5.857	2.376 2.220	0.266	0.793
PD	Low High	4.857 5.048	2.476 2.674	-0.594	0.559
TD	Low High	5.143 5.571	2.220 2.638	-1.071	0.297
PF	Low High	3.571 3.952	1.469 1.717	-2.019	0.057
EF	Low High	5.381 5.810	2.312 2.358	-1.337	0.196
FR	Low High	4.762 5.143	2.448 2.496	-1.053	0.305
TO	Low High	4.944 5.230	1.967 2.172	-1.057	0.303

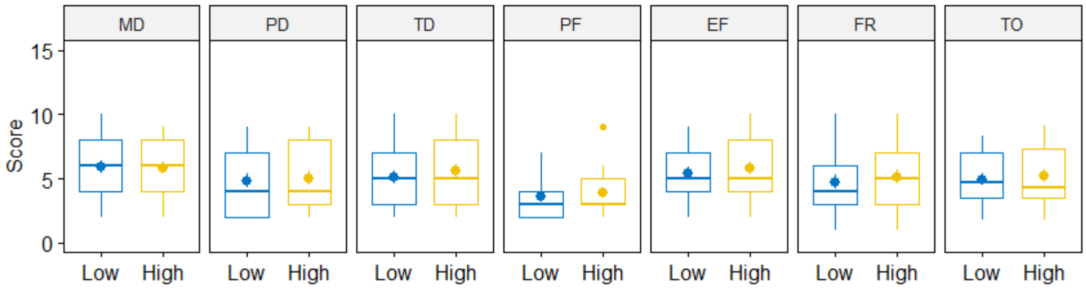


Figure 6.3: Resource management task scores between low and high-demand level. The dots within the box indicate the mean of the scores, with standard error of the mean.

In the resource management task, the results suggest that task demand levels of this particular task cannot be differentiated accurately before the task. Our results suggest that participants cannot subjectively perceive the difference between low and high-demand task when seeing a resource management task.

MULTI task. A paired-samples t-test with Bonferroni correction was conducted to compare scores from all dimensions of NASA-TLX scales in low and high-demand level of MATB multitasks, where system monitoring, tracking, and resource management tasks were put together in an integrated task (MULTI task). In these tasks, we found there was no significant difference of NASA-TLX scores when seeing low and high demand MATB multitasks (see Table 6.4 for the statistics and Figure 6.4 for the graphs).

These results suggest that task demand levels of MULTI task cannot be differentiated accurately before the task. Our results suggest that participants cannot subjectively perceive the difference between low and high-demand task when performing all MATB tasks.

Table 6.4: The results of t-tests for MULTI task.

Dimension	Level	Mean	SD	t(20) = ?	p
MD	Low High	7.762 8.333	1.700 1.461	-1.549	0.137
PD	Low High	7.048 7.619	2.269 1.936	-1.638	0.117
TD	Low High	8.143 8.286	1.590 1.736	-0.354	0.727
PF	Low High	5.429 5.429	1.912 1.912	0.000	1.000
EF	Low High	7.762 8.143	1.841 1.878	-1.073	0.296
FR	Low High	7.333 7.667	2.153 1.958	-1.128	0.273
TO	Low High	7.246 7.579	1.558 1.459	-1.224	0.235

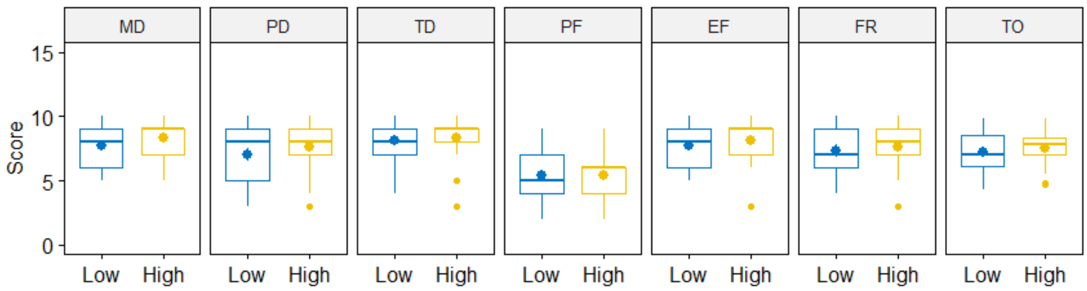


Figure 6.4: MULTI task scores between low and high-demand level. The dots within the box indicate the mean of the scores, with standard error of the mean.

6.5 Discussions

This experiment aimed to investigate whether MWL during the MATB task can be predicted prospectively. Our hypothesis stated that there will be significant difference in prospective prediction of subjective workload to MATB task, both for its individual components or as an integrated task. However, the results from this experiment only supported the hypothesis partially. We found significant difference in participants' prediction mostly in system monitoring task; while all tasks were put together, participants seemed to be unable to distinguish task demands levels. The compilation of the t-test results can be seen in Table 6.5.

The results of this experiment might suggest that participants' expectations about task demand varied between MATB subtasks. In general, as seen in corresponding tables, participants rated tasks with high demands with higher scores in all MATB subtasks including the MULTI task, with very few exceptions. Performance score in the integrated tasks of MATB was rated indifferently between low and high-demand tasks, while mental demand score was rated lower in high-demand task in the resource management task. However, significant differences were only found partially, mostly when participants were seeing system monitoring task as shown in Table 6.5. Up to this point, the results from this study have similarities with study from Sublette et al. (2009, 2010) in terms of general differences in subjective MWL scores. However, our studies appear to be dissimilar regarding the statistical results, generating different explanations.

Table 6.5: Compilation of the t-test results from the experiment.

Task	**MD**	**PD**	**TD**	**PF**	**EF**	**FR**	**TO**
SYSMON	*	**				*	*
TRACK						**	
RESMAN							
MULTI							

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

From our study, the results might suggest that the system monitoring task was the only MATB subtask whose demands were possible to be predicted by merely seeing the prospective task. From the results as well, participants could predict that the demands could result in different perceptions of mental demand, physical demand, frustration, and thus MWL score in general. Meanwhile, in the tracking task, significant difference was found in frustration score, suggesting that this was the only aspect that can be distinguished by participants. Nevertheless, when all of these subtasks were put together (MULTI), participants were unable to distinguish task demand levels at all. Therefore, we might conclude that MWL level of MATB task cannot be predicted before the actual task.

These tendencies might come from stimulus disparity of the MATB tasks that are a 'signal detection' task that require participants to respond accordingly. Everly (2016) suggests that in signal detection, task with high stimulus disparity tends to be more distinguishable. In high-demand level of system monitoring task, rapid changes of the task can be clearly seen, with green and red lights 'flip-flopping' and four individual scales going up and down swiftly. The response required for the success of this task, in real simulation of MATB, is by immediately pressing corresponding keys on the keyboard. The nature of this task might create perceptions among participants that this particular task was demanding mentally and physically, thus creating frustration when performing for real. This task, consequently, was considered generating significant mental workload.

Contrary to a system monitoring task with many attributes, the only interest in the tracking task was the target cursor that is moving out from its targeted area. The expected response of this task was to direct and hold the controller (usually the joystick) to the targeted area. This nature of the task might not create sufficient impression of mental demand or frustration as in a system monitoring task. Whilst controlling with the joystick might be physically effortful (depending on the characteristics of the joystick itself), yet participants seemed unable to 'imagine' the experience. This might be because controlling with the joystick might be uncommon for most participants. Even in the common context in life such as gaming experience, a controlling device such as the joystick might not be the main necessity. This inability to experience the physical element of the task might be related to effort and performance scores in all subtasks and MULTI task. In NASA-TLX brief explanation about the dimensions, 'effort' dimensions were described as 'how hard' the task would be and performance as 'how successful' the result of their performance would be. These two elements appears to be difficult to comprehend since they had not experienced the task in reality. Similar to the physical element, the sense of 'rushing' in responding to the task was also seemingly unimaginable if not experienced in reality. This might also explain of why temporal demand failed to be differentiated in all tasks.

Meanwhile, the resource management task requires a more complex cognitive process as participants must calculate the proper distribution of fuels across tanks, with the pumps failing occasionally during the session. Therefore, resource management seems to be the only subtask in MATB whose expected response is not immediate. Participants would have a variety of responses since the task was involving different strategies among participants; and when the strategies would be implemented was completely at participants' discretion. Seeing the video of resource management might not sufficiently create the sense of immediacy in responding to both demand levels. Moreover, as mentioned in Study 3 discussion, the resource monitoring task seems to be less-dynamic compared to the system monitoring and tracking tasks. Participants, therefore, failed to distinguish demand levels of the task in terms of all NASA-TLX dimensions, including the total score of MWL. This explanation may also support the results from experiments in Study 3 revealing that resource management scores were not different in both demand levels.

Concerning the integrated task (MULTI), it can be said that the task cannot be differentiated between both task demand levels. The possible explanation of this result was that putting all subtasks together had made the disparity of the elements of the tasks faded. In other words, the ability to discriminate these individual tasks were confounding each other. Various yet not-too-different stimuli that have to be seen in the video of MULTI task may explain the indifferent results between task demand levels. From the discussion so far, despite partially supported hypothesis (in system monitoring and tracking task), we may conclude that we cannot predict MWL during a simulated flying task prospectively. Without experiencing the task in reality, it appears to be impractical to assess how 'difficult' or 'busy' the task would be unless the elements of the task are considerably contrasting.

This study was clearly preliminary, leaving numerous questions to be investigated further such as the effect of expectations to the choice of strategies, the impact of experience and individual differences, or the consistency between prospective and retrospective judgements. We also acknowledge limitations of this study, particularly regarding the methodology and the scope. Future research might be interested to broaden the scope, for example, by undertaking this study with subject-matter experts. Linking prospective and retrospective evaluation of MWL in more proper laboratory environment might also be beneficial to check the validity and reliability of the expectation. With its preliminary nature and limitations, however, the study has provided valuable information in understanding MWL in the context of a flying task.

6.6 Chapter Summary

This last empirical study aimed to understand if mental workload during a simulated flying task can be predicted pre-task. An online experiment was undertaken to demonstrate the ability of participants in distinguishing task demand and thus MWL by merely seeing the videos of the tasks. The results suggested that, in general, predicting a simulated flying task prospectively appears to be impractical, despite little evidence supporting the notion, particularly in the system monitoring and tracking tasks. This preliminary study provided a different angle in understanding mental workload specifically in the context of a flying task. The next chapter will discuss all the lesson-learned together and, ultimately, conclude the thesis.

Chapter 7

General Discussions and Conclusions

7.1 Chapter Overview

This chapter concludes the thesis by discussing general findings from the research and their contributions to broader discourse of relevant research topics. The organisation of this chapter consists of four parts, namely (1) revisiting the research aims, the way they were achieved through a series of studies, and the main findings along with their relations to each aim; (2) discussing the validation progress for the research and contributions resulting from the research in the sense of their attempts to fill the gap in the literature; (3) addressing issues regarding the approach and methods applied in the research, particularly the efficacy of physiological and cerebral measurement in predicting mental workload; and, finally, (4) acknowledging limitations of the research while suggesting directions for future research on similar topics and implementation/explorations resulted from this thesis. Concluding statements at the end of this chapter will provide conclusions of this thesis.

7.2 Summary and Reflections on Research Aims

7.2.1 Study 1: Pilot and Public Acceptance on MWL Sensors

This study initiated the journey of understanding MWL in the context of flying. Within the literature, attempts to measure MWL using more objective indicators have been conducted. However, acceptance from pilots and the public towards this idea was seemingly less-explored. This study, therefore, aimed to explore attitudes from professionals and the public towards the implementation of MWL sensors. To achieve the aim, we surveyed pilots and conducted a more in-depth analysis using Focus Group Discussion (FGD) interviews. Furthermore, an online experiment involving the public was also undertaken.

Our results suggested that most pilots we surveyed agreed with the implementation of MWL sensor technology in the future. Using the TAM framework (Davis, 1989), we analysed the way this attitude was formed among pilots. The results implied that if the technology was perceived to be 'easy to use', perceptions of usefulness would be formed and, both directly or indirectly through positive evaluation, this could lead to an intention to use the technology. Deepening and supporting the responses from some participants, FGD interviews found both positive and negative attitudes towards the implementation of the technology. The sensors were believed to help pilots in developing awareness about their MWL. Pilots tend to ignore their 'state of busyness' due to the urgency to complete a flight mission. With this tendency, measurement of MWL in a real-time manner, particularly using a more objective indicator, is considered important to remind them if they are 'too busy'. This arguably aimed to create standardised metrics of MWL that can be accepted and understood uniformly, and it may prevent resistance from pilots to change strategies in completing the flying task. However, the key issues found were related to whether the technology would be accurate enough to detect MWL, whether human pilots would still have authority to control the aircraft, whether the sensors would be comfortable to wear, and whether the data collected from the sensors would be securely stored. As professionals seemed to agree, members of the public would also agree with the implementation of such technology. From the experiment, we found that members of the public would be more confident to fly with pilots who are using MWL sensor technology.

7.2.2 Study 2: The Qualitative Approach to Identify Sources of MWL during Flying

From Study 1, we have understood that measuring MWL using hypothetical physiological sensors could be accepted by pilots and the public. Before demonstrating the ability of the sensors in detecting MWL during a flight, understanding the nature of the flying task was essential, particularly from a cognitive standpoint. Hence, this study aimed to investigate cognitive processes involved when performing a certain phase of a flight, in this case, landing. By understanding cognitive processes during a real flying task, MWL of pilots can be understood as well. To achieve the goal, a Critical Decision Methods (CDM) interview was undertaken with five professional pilots. The results of this study identified sources of MWL related to demands, performance, and external/internal influences. Six factors influencing MWL during a real flying task (landing), including data preparation, ATC communication, plan updating, outside monitoring, aircraft configuration, and manual flying. These demands generated MWL, and MWL subsequently generated performance feedback that can be seen from display/auditory feedback, physical changes in the aircraft, and from the flying partner. External and internal influences on the pilots, particularly from experience or manuals from manufacturer and company, also affected task demands in the sense that they may add more demands or, oppositely, decrease them.

7.2.3 Study 3: The Objective Approach to Predict Pilots' MWL

From Study 2, we have comprehended that MWL during a flying task was dynamic instead of constant. Consequently, potential MWL sensors have to be tested in an environment that can resemble the real context of flying. Study 3, therefore, aimed to demonstrate whether MWL changes during a simulated flying task can be predicted using more objective indicators. Two experiments have been undertaken to achieve the goal. The first experiment investigated the relationship between fNIRS and MWL, while the second experiment utilised other physiological measures (heart rate and pupil dilation sensors) to examine their relationships with MWL. The results from these two experiments were discussed in two manners: (1) whether the individual sensor can observe changes in MWL and (2) whether the individual sensors correlate with subjective measures of MWL.

It was suggested from Experiment 1 that fNIRS seemed to be unable to distinguish the low and high-demand task. We evaluated the methodology and techniques from

Experiment 1 and, with few changes in Experiment 2, the results suggested that fNIRS responses could be seen differently in low and high-demand task in all eight channels, with significance was only found in Channel 6 and 7. However, if we considered brain activation in all channels from their respective baseline (zero), the high-demand task did activate the whole measured region of the brain; while the low demand task only activated Channel 1 and 2 that corresponds to the right side of prefrontal cortex. In Experiment 2 as well, heart rate/variability (RMSSD) and pupil dilation supported previous research suggesting these physiological changes could predict MWL changes. Regarding fNIRS, our results indicated that fNIRS were able to demonstrate brain activation changes, yet it seems insufficiently sensitive for such a simulated flying task. While the direction of correlations with subjective measure of MWL were observed to be as expected (except in Channel 5 fNIRS in Experiment 2), with various degree of correlation coefficient, statistical significant correlations were only partially found in Experiment 2 (Channel 2 fNIRS and RMSSD). In both experiments, the role of anxiety and motivation to the physiological measures of MWL were not found.

7.2.4 Study 4: Vicarious Observation of MWL

Studies 1 to 3 provided insights about the MWL during a flying task using both qualitative and quantitative approaches with various techniques such as survey, interview, and physiological measurements. These approaches, however, predicted MWL after or during the task. Since operators might have expectations regarding the task they are going to perform, understanding this could extend our understanding of MWL during a flying task. Study 4, therefore, aimed to investigate whether MWL of a task can be distinguished before the actual task. An online experiment using videos of the MATB task was undertaken to achieve the aim. This study was considered a preliminary attempt in exploring this area. The results from this experiment suggested that task demands of MATB cannot be distinguished before the task. Some individual tasks of MATB, the system monitoring and the tracking task, can be distinguished in partial dimensions of subjective MWL measure (NASA-TLX).

7.3 Contributions Resulting from This Thesis

The research presented in this thesis has contributed to the wider discussion of the MWL topic in the manners as shown in Figure 7.1.

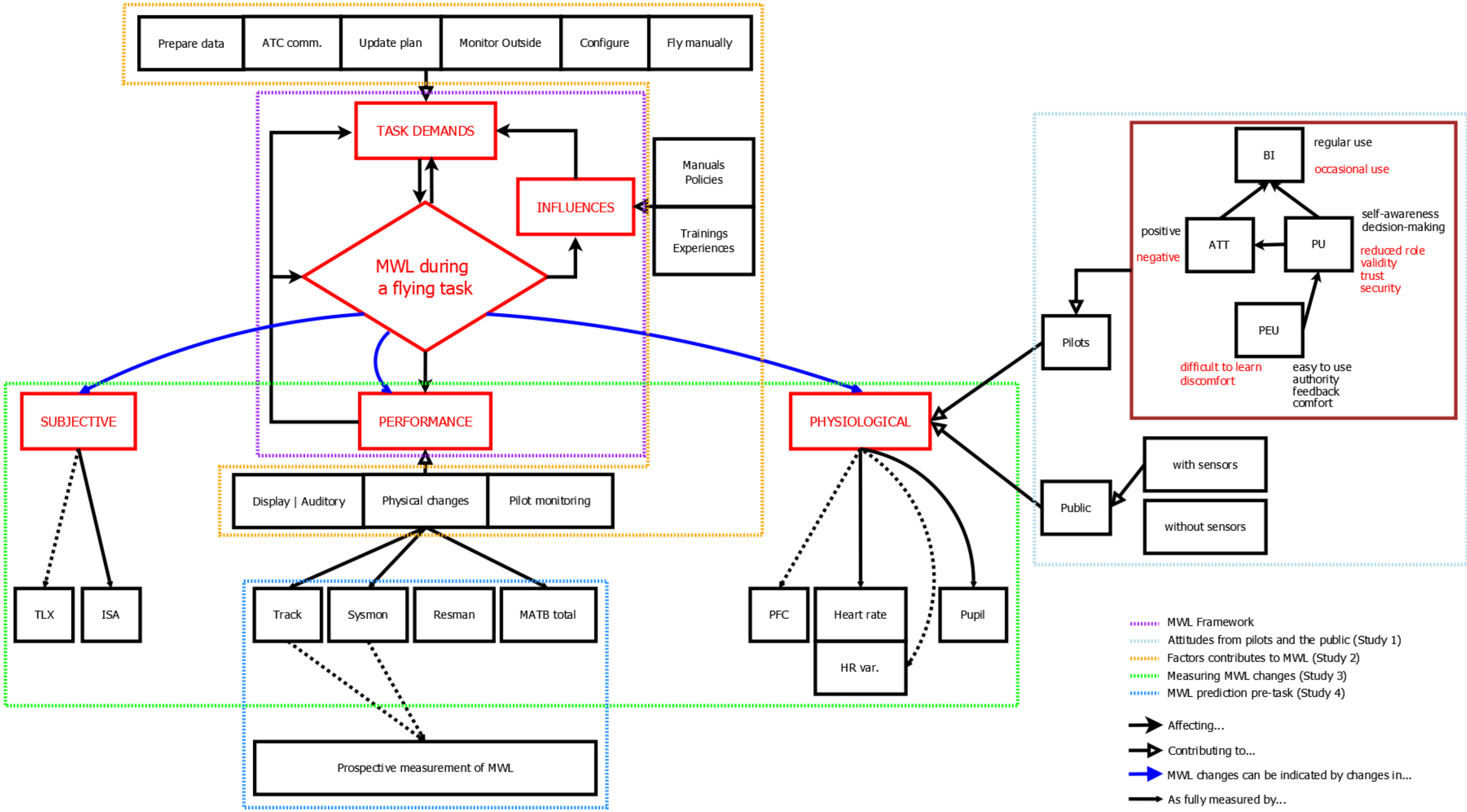


Figure 7.1: The graphical representation of the research contributing to the proposed thesis.

We summarise our thesis as depicted in the figure as follows. The MWL measurement framework from Sharples & Megaw (2015) was used to understand the formation of MWL during a flying task, i.e. the landing phase. This framework is shown in an area of Figure 7.1 surrounded by the purple dotted line. According to this framework, the MWL was formed by the interaction of task demands, influences from inside or outside the system, and performance feedback. Study 2 was directly linked to this framework as this study had successfully identified the sources of each MWL components and their interaction during the landing phase (shown in an area of Figure 7.1 surrounded by the amber dotted line). The results of this study suggested that the physical/mental demand came from data preparation, ATC communication, plan updating, outside monitoring, aircraft configuring, and manual flying. Meanwhile, the performance feedback came from the instruments, physical movement of the aircraft, and pilot monitoring feedback. Manual and policies, as well as experiences and training, formed internal/external influences component.

The MWL changes during a flying task was then demonstrated using three different measurements, namely subjective measures, performance indices, and physiological indicators. However, we demonstrated this using a simulated flying task (MATB). From Study 2, one of the sources for performance feedback was physical changes of the aircraft. They were simulated in MATB tasks, such as the tracking task, the system monitoring task, the resource management task, and the combination of these individual tasks. Along with subjective dan physiological measurements, this performance indicator captured changes in mental workload during a MATB task as presented in Study 3 (shown in an area of Figure 7.1 surrounded by the green dotted line). More specifically, differences in the MWL were detected by ISA scale as well as by the tracking, the system monitoring, and the combined MATB tasks scores. Moreover, changes in heart rate and pupillary dilation could also indicate MWL changes. Meanwhile, in addition to NASA-TLX scores, some measurements such as differences in the PFC oxygenation and in heart rate variability could only detect MWL changes partially.

Study 3 led to an exploration to prospective MWL measurements as presented in Study 4, which resulted in limited information about prospective MWL prediction (shown in an area of Figure 7.1 surrounded by the blue dotted line). Finally, Study 1 supported the explanation for physiological measurements of the MWL in the context of aviation. This study provided insights about professionals' and the public attitudes towards the implementation of MWL sensors technology in the future (shown in an area of Figure 7.1 surrounded by the greyish-blue dotted line). The public would prefer flying with pilots whose MWL are monitored. Meanwhile, the pilots supported the implementation of MWL sensors in the cockpit through attitude formation based on their perception on the ease of use and the usefulness of the technology.

The more detail explanations of the contributions from each study will be discussed in the following subsections.

Contributions of Study 1. The first empirical chapter of this thesis (Chapter 3) contributes to preliminary investigation of professionals' and the public attitudes assuming MWL sensors technology are about to be integrated in cockpit system in the future. The technology acceptance study in this chapter provides insights into the attitudes towards the implementation of MWL sensor technology in general and the way these attitudes are shaped among professional pilots. In general, the generic model of TAM did not agree with our data, particularly regarding the perceived ease of use (PEU) variable that cannot directly cause the intention to actually use the technology. The variable had to be mediated by the way pilots perceive its usefulness or with addition to attitude mediation. This result was similar to the result from Richardson et al. (2019). In their study, the ease of use has strong positive correlation with perceived usefulness. This perception, therefore, leads to behavioural intention to use the technology. However, similar to ours, the results from their study did not support the relationship between perception of ease of use and behavioural intention.

This study also provides insights from the public as one of the main stakeholders of aviation that would be indirectly impacted by the implementation of the technology. Since the exploration in this particular topic is seemingly still sparse, we cannot compare our results with previous reports. However, studies on public perception regarding autonomous airliner (Vance & Malik, 2015) and single-pilot operation airlines (Stewart & Harris, 2019) could bring some insights. Both studies suggested the disagreement of the public to the implementation of pilotless or SPO aircraft. We interpreted these results that the presence of pilots is still essential to the passengers for safety reasons. Meanwhile, the implementation of MWL sensors will not replace the pilots but, in fact, it could improve their reliability. Therefore, this can be perceived by the public as an effort to increase safety.

Using two difference theoretical frameworks, the conclusions seem to converge, that both pilots and the public show positive attitudes towards the technology implementation. This means that pilots would use the technology during flying and members of the public would fly more confidently if their pilots used the technology. Several concerns, however, have also arisen from this hypothetical technology from pilots, such as security or validity issues in particular. All of these opinions, including concerns arisen, could serve as a guidance in designing the system once it has been ready to be implemented in the future.

The results from this study have contributed to the thesis by revealing that MWL comprises subjective and physical experience of the task. Pilots supported the notion that MWL needs to be measured more objectively, since it was often mainly seen as a subjective experience. However, the implementation of such technology may increase their workload if the sensors were not carefully designed, since it may add unnecessary physical burden during a flying task; this may create another workload. Public opinions also supported the notion that MWL tends to be subjective. Therefore, the implementations of MWL sensors may increase pilots' reliability and subsequently flight safety. All of these results could inform us that the implementation of MWL sensors technology would be perceived as a positive endeavour to improve aviation safety from the perspective of MWL management.

Contributions of Study 2. The CDM interview method, as used in the second empirical chapter of this thesis (Chapter 4), is commonly used for eliciting information on cognitive functions during tasks (Crandall et al., 2006) and does not directly aim to measure mental workload. Since mental workload emerges from interaction of several cognitive activities, however, we utilised this method to predict mental workload by tracing it retrospectively during actual tasks. To our best knowledge, this study was novel as it might be the first utilisation of a CDM interview for predicting mental workload of pilots qualitatively. Previous studies use CDM for cognitive analysis during naturalistic decision-making (Cattermole et al., 2016), during in-field collaboration among traffic incident responders (Cattermole-Terzic & Horberry, 2020), or when responding in-flight power-plant system malfunction (PSM) (Asmayawati & Nixon, 2020).

From this study, we tend to support the notion that MWL involves subjective elements to these interactions (e.g. Van Acker et al., 2018); and the degree of involvement appears to be substantial. The way pilots think and decide during landing are influenced by their previous experiences. This part of the thesis therefore contributes to the understanding of pilots' cognitive activities and their role in generating mental workload during landing phases. The method used in this study was also helpful to gather more insights as it was based on pilots' actual experiences instead of merely general knowledge or procedures about landing. The results from the CDM study also revealed that pilots' jobs consist of various independent tasks that run continuously throughout the landing phase. This information is useful to create stimuli that could resemble the real-life job context in a controlled laboratory experiment (i.e. the brain and physiological study in Chapter 5).

In this study, we used the MWL measurement framework from Sharples & Megaw (2015). For the context of our qualitative study, the framework was useful to limit possible sources of MWL. Using the framework, we were able to explore more deeply about particular operational tasks that may influence each MWL source. However, we found that the framework seems to be oversimplified. The framework does not specify the inside processes of task demands forming, for example, by associating it to specific MWL theory such as the MRT (Wickens, 2008). The specific measurement of task-switching (Schmitz & Voss, 2014) may not be accommodated by the framework. Whilst this framework is useful for measuring MWL in a more practical environment or contexts, this framework may not be suitable for more detail MWL measurement.

Using a qualitative approach, the study has contributed to the thesis in terms of providing understanding about sources of MWL and their interaction during a flying task. From professionals' minds revealed on this study, we could be informed that the tasks pilots must undertake were considerably dynamic due to combinations of subjective and 'objective' elements of MWL. Task demands during a flying task might be seen as a quantitative element, such as bad *vs* good weather, and thus could generate different MWL. However, these demands are also influenced by pilots' subjective assessment that comes from experience. Experience plays an important role in this matter, as it helped pilots to create better judgment for decision-making and thus better performance. Even though the step-by-step procedures for landing tend to be similar, particularly if the aircraft was made by the same manufacturer, the way pilots respond to events during this crucial phase was different. The results from this study may have implications to the understanding of MWL, particularly during a real flight operation. Interaction between demands, performance, and internal/external generated MWL dynamically. Changes could occur unpredictably, thus putting MWL in constant change. For example, abrupt change of ATC instructions might generate higher MWL relative to the initial MWL that was occurring.

Contributions of Study 3. Chapter 5 of this thesis (Chapter 5), or the brain and physiological study of estimating MWL, contributes to our further understanding of the use of fNIRS, heart rate, and pupil dilation in capturing MWL changes during continuous, long tasks. From our series of experiments, we confirm that fNIRS seems insufficiently sensitive to detecting MWL changes in a simulated flying. As shown in our second attempt of the experiment, fNIRS can show activation in all channels when participants were exposed to a high-demand task, and in few channels when exposed to a low demand task. Statistically speaking, however, the ability to differentiate between high and low demand tasks can only be shown in few channels (Channel 6 and 7 in our case). The results support previous studies such as Ayaz et al. (2012)

that successfully demonstrate fNIRS ability to detect MWL changes in general and specifically in the edge (left and right) regions of prefrontal cortex (e.g. Causse et al., 2017; Galoyan et al., 2021).

Still, our results also tend to support the notion of fNIRS insensitivity to certain types of tasks, as in a study from Argyle et al. (2021). Along with brain activation measures, we applied heart rate and pupil dilation measures to cross-validate the measurements. As mentioned earlier, our findings seem to agree with previous studies using heart rate, heart rate variability (RMSSD) (Thayer & Lane, 2009), and pupil dilation (Eckstein et al., 2017) in terms of the way these underlying organs behaved in responding MWL changes. Nevertheless, the association between these brain and physiological data and overall subjective experience of MWL needs to be investigated further. The correlations mostly show correct directions as theorised, but the quality of these correlations seems to show inadequacy. In general, the results from this study support results from previous studies, particularly about the MWL measurement using brain and physiological indicators. However, from our studies, some measures have shown insensitivity, possibly due to the nature of the task that was too complex.

The results from Study 3 contributed to the thesis by providing understanding of MWL from the perspective of more objective measurements. Subjective experience of MWL were apparently related to the certain physiological responses, thus making physiological measurements of MWL feasible. However, the extent to which the sensors can detect MWL changes might depend on the task. Task environment with dynamic stimuli such as MATB might not be clearly distinguishable for fNIRS in particular. The results may deliver implications to the understanding of MWL in a real flying context. The detection of MWL changes by certain physiological sensors during a real flying task could be very slight. This might be due to the task itself that was too dynamic or too varied in its demands to be measured easily by fNIRS. Further investigations might need to be done before implementing this MWL detection method in a real flying environment.

Contributions of Study 4. The last empirical chapter (Chapter 6) attempts to investigate whether MWL during a flying task can be predicted before the actual task. From our online experiment, participants' expectations about task demand varied between MATB subtasks. In general, participants rated tasks with high demands with higher scores in all MATB subtasks including the integrated task, with very few exceptions. However, significant differences were only found partially, mostly when participants were seeing SYSMON tasks. Up to this point, the results from this study have similarities with study from Sublette et al. (2009, 2010) in terms of general differences in subjective MWL scores. The conclusion drawn from this experiment is that we cannot predict MWL during a simulated flying task prospectively. Without experiencing the task in real-life or in simulation, it appears to be impractical to assess how ‘difficult’ or ‘busy’ the task would be unless the elements of the task are considerably contrasting.

The results from Study 4 contributed to the thesis by providing an explanation of the inability to predict MWL prospectively during a flying task. Whilst task demands in some individual tasks might be distinguishable without experiencing them in reality, the entire flying task has to be experienced in actuality to obtain accurate prediction of its MWL. The results from this study might have implications in the real practice of flying. More specifically, it appears to be impractical to tell pilots how ‘busy’ or ‘difficult’ a flight would be by merely providing description of the entire flight conditions. It might be possible, however, to tell them certain aspects of the flight that have understandably contrasting elements, such as weather conditions. Since this study was a preliminary, further investigations are needed to broaden our understanding of prospective prediction of MWL.

Summary of Contributions. Based on the discussion presented previously, the summary of contributions resulting from this thesis is therefore presented in the following points:

1. This thesis contributes to provide insights about professionals’ and public attitude towards MWL sensors technology in the future. The study also revealed concerns among professionals and members of the public around this kind of technology. This analysis informs decision makers and technology designers when realising this technology in the future.

2. This thesis contributes to the further understanding of what makes pilots experience high mental workload during landing. It has been agreed that landing is the most crucial and demanding part of a flight due to the complexity of the task, requiring pilots to allocate all of their cognitive resources. However, what pilots actually think and do could be very dynamic depending on the situation. Five CDM interviews revealed sources of the demands and their interaction during a landing procedure.

3. This thesis contributes to the measurement of mental workload using physiological methods. Regarding fNIRS, our study has contributed to supporting the spatial resolution of the fNIRS. More specifically, our results suggest that the left-side of the prefrontal cortex was activated in response to MWL changes

during the simulated flying task. The notion that heart rate measures and pupil dilation can indicate MWL changes were also supported by this study.

4. This thesis contributes to the initial understanding of prospective MWL prediction during a flying task. Our work offers a notion that MWL during a simulated flying task cannot be accurately predicted before the task, unless the contrasting elements of the task can be shown. This study may also be seen as a support to the mainstream method of MWL measurement, i.e. that MWL changes can be accurately captured during and after the task.

7.4 Validation of the Research

The process of validation for the research was started by establishing external validity through literature review. Once we had obtained initial support for the research concept, internal validity of the concept was then supported by findings from the studies. In terms of internal validity, we started to establish the concept of potential implementation of more objective measurement of MWL using physiological methods and gather opinions from professionals and members of the public about the concept. After obtaining insights about pilots' and public attitudes, we attempted to test the concept of more objective MWL measurement. However, before performing the test, we first confirmed the proposed simulated flying task with the real experience of pilots when flying an aircraft. This aimed to validate the simulated flying task so that we can be assured that it resembled the real flying task in terms of cognitive functions being exercised. This validation also applied to prospective prediction of MWL during a flying task.

7.5 Reflections on Approach and Methods

Throughout the research, we have applied various methods and techniques in our studies. In this section, we will discuss approaches and methods used in this thesis and what we can learn from the experience. For pragmatic reasons, we group the discussion into two main topics, namely (1) physiological measurements and (2) qualitative and online studies.

7.5.1 Lessons Learned from Physiological Measurements

Stimuli and procedures for fNIRS. From the experiments in Study 3 (Chapter 5), we found two different outputs as the results of two different procedures. Since we wanted to produce stimuli that resemble flying tasks as realistically as possible, five-minute MATB tasks were used as they were considered appropriate to evoke MWL changes as in real-life setting. From Experiment 1, we arranged high and low demand tasks in a row without any interruptions or rest period, except a 20-second period of NASA-TLX administration after each five-minute stimuli. The results showed increase on the first five-minute block regardless of the manipulation (high- or low-demand tasks) and followed by lower activation pattern of brain activation. Figure 7.2 shows the evidence.

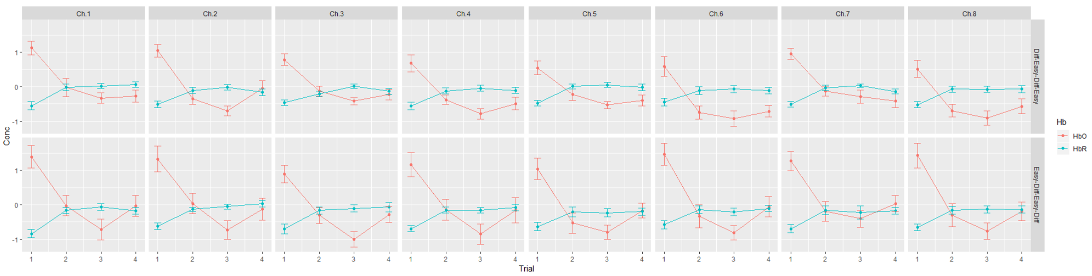


Figure 7.2: The results from the first experiment of the brain and physiological study (Chapter 5) on each block.

We evaluated our procedures and found that fNIRS signals would work better to distinguish between high and low demand stimuli if given some time to return to baseline after each block. As we mentioned in Chapter 5, there is no agreed period of time for rest or baseline, but it is suggested to have a baseline period that comes near to the period of the stimuli (Herold et al., 2018). Regarding Experiment 1, we thought that NASA-TLX administration period could function as baseline period, even if it was set short (around 20 seconds). Instruction manuals and prior training in using the fNIRS device suggested various baseline periods; a baseline period that lasts for as short as 10 seconds would work in most cases. However, as there was a steep increase in every first block of the stimuli, we concluded that our tasks might be too complicated to allow signals to return to baseline within such short period of time. It was also suggested that completing the NASA-TLX may have involved some degree of cognitive demand; thus this cognitive activity may inadvertently get picked up by fNIRS during the baseline period. Therefore, an extension of baseline period, separated from NASA-TLX administration, may produce better results. A one-minute baseline period was then applied in Experiment 2. This decision was made with several considerations, i.e. finding an optimum equilibrium between coming near to stimuli period and avoiding possible mind-wandering due to baseline period that lasting too long (Herold et al., 2018). As the results, in Experiment 2 we could see brain activation (from baseline/zero) in response to the stimuli and few channels could even differentiate between high and low demand stimuli. We thus conclude that the decision in choosing and presenting the stimuli plays essential roles in fNIRS data output.

The training of participants to get familiarised with the task was also important. In our first experiment, we only introduced participants to basic features of MATB and allowed participants to experience the tasks for short periods of time. We suspected that this might contribute to the results of Experiment 1. Learning from this situation, and also following suggestions from Jaquess et al. (2018) and Ayaz et al. (2012) regarding participants training for a simulator study, we introduced a full version of the MATB tasks (either low or high-demand conditions) for training participants. According to these authors, the combination of exposed time and repetitions of the tasks during the training would be sufficient to guarantee that they would understand the tasks, thus reducing potential trivial behaviour when completing the tasks.

Different signal preprocessing, different result. fNIRS data cannot be analysed directly as we need to clean the signals from possible confounding variables or artefacts such as heart beat or participants' movement. Various pipelines of signal preprocessing have been proposed by previous researchers. However, signal preprocessing pipelines seem to be idiosyncratic. In simpler words, different pipelines would produce different outputs; this might subsequently affect the analyses. These phenomena can be seen from previous studies that tend to use different preprocessing pipelines. Some studies even attempted to compare pipelines and recommend which one would fit for a specific study (e.g. Pinti et al., 2019; Cooper et al., 2012). Generally, there are various 'menus' of signal preprocessing methods. In the software we used throughout the studies (Homer2), there are at least 24 functions to perform specific task in cleaning the signals (Huppert et al., 2009). fNIRS researchers then have to combine the functions and set parameters on each functions to obtain cleaner signals. If they believe their signals are too confounded due to, for example, extreme movement of the participant, they need to choose motion artefact correction functions and perhaps with radical parameters. However, this strategy might increase chances of losing too many signals. fNIRS researchers therefore need to choose functions and set parameters according to their necessities and prepare sufficient judgment of them. The main aim of signal preprocessing, in layman's term, would be to obtain signals that are 'good enough' to be processed without losing much of them.

Having been aware of the phenomena, we attempted to compare our chosen pipeline (details in Chapter 5) with another pipeline from Di Lorenzo et al. (2019). The results suggest that our pipeline was less conservative, whereas the alternative filtered out too many signals (see Figure 7.3). From visual inspection of sample signals, we concluded that our signals were ‘good enough’ to be processed, as motion artefacts seemed to be minimal (see Figure 5.6). This signal preprocessing experience provides an insight that justification for signals have to be made from both objective and subjective methods. Objective methods come from set of parameters that are believed to be proven in previous research and also ideal for our research, whilst subjective ones can be done by visually inspecting the sample signals. However, some biases might appear when applying subjective inspection.

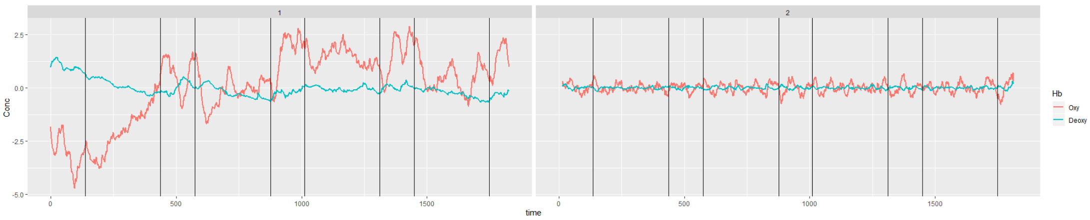


Figure 7.3: Corrected signals from our pipeline (left) and pipeline from Di Lorenzo et al. (2019) (right).

Signal filtering issues also emerge from heart rate measurement. We applied different level of correction from ‘very low’ to ‘very strong’ (as in Kubios HRV Standard software) for each block and for each participant. We started from ‘very strong’ methods to guarantee as minimum artefacts as possible (less than five percent as suggested in the manual). If the correction results in more than five percent artefacts, we lowered the level to ‘strong’. This process went on until we found less than five percent of artefacts. If this process failed even after applying ‘very low’ level of correction, we decided to exclude the data as it suggested that the signals were too noisy. With this method, one participant might lose data from one or more blocks of the experiment. Since the experiment consisted of two blocks of high and low demand task respectively, we set a rule for data inclusion that minimum one block data must be preserved for each demand level. If one participant lost data from two blocks of, for example, high-demand tasks, this participant was excluded since data comparison would not be possible. This particular phenomenon might be sourced from inconsistent sensor placement. Details for this will be discussed in the next section.

Sensor placement affects the quality of data. Sensor placement may affect the quality of the data. We started the discussion of this issue from heart rate measurement. As mentioned in the previous section, the way participants placed the heart rate sensors in their body was uncontrollable due to privacy issues. For this experiment, we asked participants to fit the sensor (and the strap) independently in a restroom after being given the instructions. To make the task easier, we provided a picture showing them where the sensor must be located, and practically showed them the ideal tension; thus the sensor would fit their bodies properly, i.e. not too tight that could cause discomfort but not too loose that could cause error in reading and displacement. For male participants, we checked directly the strap placement and tension if allowed for quality control. With all of these efforts, we believed to reduce chance of errors in reading heart rate signals by the sensor. Nevertheless, errors might still possibly come from a small fraction of participants. Acknowledging this issue, we chose to initiate the signal correction method using the most conservative parameter.

Contrary to heart rate measurement, issues regarding fNIRS sensors did not come from privacy-related matters. Instead, they tend to be technical, i.e. differences in the size of forehead areas. The fNIRS device we used has fixed or pre-defined location of sensors or optodes from the manufacturer. This may therefore affect the accuracy of the areas under investigation as the outermost channels such as Channel 1, 2, 7, or 8 might be eventually sitting slightly off from the prefrontal cortex for participants with considerably narrow foreheads. The sensors will be possibly capturing activity from motor cortex (Brodmann Areas of 6 or 4) instead of prefrontal cortex (Brodmann Areas 9 or 10). Moreover, in these particular participants, the sensor's contact to skin would be slightly obstructed by hairs such as eyebrows or side hairs. The effect of obstruction, in several cases, can be seen directly from live signals that appeared in the software, either shown by moving back-and-forth vertically or, oppositely, complete flat indicating 'no activity'. A similar problem, based on our experience, also emerged from female participants wearing a veil/hijab (as in Muslim women) as the sensors might be obstructed by the fabric. For this particular case, we attempted to move these obstructions as much as possible by, for example, removing hairs from sensors or putting the sensors strap inside the veil. We then made sure that the signals shown in the software were 'good enough' to continue.

Another technical issue was related to skin colour. As theorised, fNIRS is sensitive to skin colour, and thus the measurements may be slightly inaccurate for individuals with dark skin due to increased light absorption by the darker skin (Wassenaar & Van den Brand, 2005). This phenomenon may cause infrared light to be unable to reach the correct depth of the brain tissue. We unfortunately did not control this issue, as our fNIRS device does not have a feature of controlling infrared light intensity. If, after sensor placement, live signals showed 'normal' attitudes we continued the experiment as we believed that the skin colour effect might be irrelevant. Signals that seem to be too noisy would be filtered out in the preprocessing stages.

Similar rules were applied to the eye tracking device. We calibrated the eye tracker for each participant before running the experiment and as long as the software indicated that the calibration was successful, we were confident to commence the experiment. However, issues might arise from participants wearing correction eyeglasses, as the light travel to the eyes could be possibly deviated by the lens. This may explain why warning for re-calibrating the eye tracker seemed to be common in participants wearing eyeglasses. For these specific participants, we re-calibrated the eye tracker before starting subsequent experiment blocks.

7.5.2 Lessons Learned from Qualitative and Online Studies

Potential biases pertaining from CDM and FGD interviews. Beside physiological measurements, qualitative approaches were also applied for the research. In Chapter 3, we used a FGD interview framework; whilst in Chapter 4, CDM interviewing was conducted. The use of CDM to elicit knowledge from subject-matter experts (SMEs) has been demonstrated in many contexts. The results from this method are substantially rich, providing researcher valuable information to create 'stories'. We tried to use this method for a slightly different purpose. Our interest was to reveal the possible sources of mental workload and the way they interact to each other. A mental workload framework Sharples & Megaw (2015) was used to guide our analyses, and thus we may bring rich data obtained from the method specifically to understand mental workload. Up to this point, we conclude that CDM is seemingly versatile in the sense of that it could be used for any purpose. The key stage of this method, from our perspective, is the analysis. We analysed the data from the perspective of mental workload, so we could further understand how mental workload emerges specifically from the interaction of demands, performance, and other factors.

Despite advantageous features, the retrospective nature of this technique might produce disadvantages as well. Hindsight bias, for example, might confound participants' accounts when discussing events under question. Hindsight bias can be simply explained as the belief that an event is more predictable after it happens than before it happened (Roese & Vohs, 2012); and these phenomena are believed to be occurring in many contexts involving decision-making (Kahneman et al., 2022). The bias is often confused with 'learning from experience' results and can produce error in decision-making when evaluating actions in the past (Roese & Vohs, 2012). Further, Roese & Vohs (2012) explained that hindsight bias is fuelled by motivational factors, i.e. an urge to see the world as ordered and predictable, and a desire to protect one's selfesteem. This issue could be relevant to our CDM application after one pilot stated about the possible influence of egocentric personality in decision-making among pilots. That is, pilots tend to consider themselves as a high-profile individual since they possess a special job that requires them to go through a difficult process of selection, training, and competition. Therefore, stories created by pilots as in our CDM interview might be confounded by the tendency to showcase their expertise in a more positive manner. We nevertheless did not specifically address this issue in the study, as this extends beyond the scope of this research. A list of questions to guide the researcher during a CDM interview, along with manual probing of their responses if necessary, could hopefully reduce this type of bias.

The aforementioned issue might also emerge in our study involving FGD interviews of pilots. The FGD interviews were conducted to gain deeper insight into pilot responses to the acceptance questionnaire. The bias that seems to emerge in a group interview such as FGD was the tendency to agree with the group. With a maximum of three pilots on each FGD session, we found that the answers from pilots in the group speaking later tended to agree with the pilot speaking earlier. If one of them wanted to convey disagreement, they tended to deliver it softly after appreciating the opposing statements from another pilot. We, however, did not measure this potential bias as it was beyond our scope, yet we found this tendency might hinder genuine response from pilots who spoke in latter order. The solution for this particular problem, in our study, was by alternating the order of speaking for each topic or question liberally. The researcher was also actively encouraging participants to communicate their opinion if necessary, for example, when one participant was found to be passive. This strategy aimed to create equal portion of opinions, thus maximising chances of obtaining rich information from all participants.

Advantages and disadvantages of online studies. Due to pandemic restrictions, we applied online studies for the research. Both the CDM and FGD interview (Chapter 3 and 4, respectively) were conducted online since we were not allowed to meet in-person. The questionnaire of pilot acceptance (Chapter 3) was also conducted online, along with online experiments to reveal public acceptance (Chapter 3) and to investigate prospective prediction of MWL (Chapter 6). Table 7.1 shows which studies in this research that were conducted online.

Even though online study is not uncommon in research, at the time of writing, current events have encouraged additional use of online or remote methods. It might be said that online studies, such as an online interview, have recently become ‘top of mind’ when choosing research methods. We evaluated the way we undertook online interviews and concluded that this method has both advantages and disadvantages. The advantage of doing interviews online is mainly about convenience of time, cost, and reachability (Wright, 2005). Working with busy pilots from different airlines and living in various cities and nations required flexibility in scheduling the session; this would be impractical if the session was conducted offline, particularly for FGD that puts a maximum of three pilots together in a single session. Consequently, time, effort, and monetary expenses associated with the session could be significantly reduced. However, the disadvantages of performing online interview have to be acknowledged as well. From our experience, technical issues (such as connection instability), motivation losses, and hindered emotional expressions are the examples of issues that might reduce the quality of the CDM and FGD interview. These considerations have to be considered in undertaking an online study.

Table 7.1: Studies using online and in-person methods in this research.

Study 1	**Study 2**	**Study 3**	**Study 4**
Survey (O)	CDM (O)	fNIRS 1 (P)	Vicarious (O)
FGD (O)		fNIRS 2 (P)	
Public (O)			

(O) = online study; (P) = in-person study

Regarding the survey and the experiment in the acceptance study (Chapter 3). The convenience rests on the ability to reach potential participants more widely. With a single hyperlink to the online session of the survey or the experiment, we may be able to advertise our study invitation to pilots, students, or public communities. This method is inexpensive and less effortful compared to conventional in-person method. However, as in offline studies, we might not be able to predict and control response rate of the invitation. That is, how many pilots or members of the public who eventually take the survey or join the experiment is still the matter of chance. In our case, 21 pilots participated in the survey and 57 members of the public joined the experiment. We believe these numbers have been an optimum result for us, given that the situation when the studies were conducted was full of uncertainty.

Concerning the experiments, along with the ability to reach potential participants more widely, the randomisation process was made easier in online experiments. The experiment in Chapter 3 in particular required participants to be allocated in two different groups with equal chances. Online experimentation made it possible to automatically put participants in either group immediately after joining the session. The disadvantage we observed regarding online experimentation was an inability to control the experiment protocol. The experiment mentioned in Chapter 6, for example, required participants to use a laptop or desktop PC since it was mainly a visual task.

However, since the online environment under which the experiment was conducted cannot be restricted to certain types of devices, it was possible that some participants accessed the session using their smartphones or tablet. This may provide a suboptimal experience for the experiment that was mainly visual, and thus affect the results.

The use of a hypothetical scenario. While the CDM study used real-world experience in the past, the acceptance study (Chapter 6) exercised a degree of participant imagination as they were asked to answer questions based on a hypothetical scenario. This means that the technology-in-question does not yet exist; hence they have no experience towards such technology. However, in the scenario in our study, we provided participants with a rough design concept of pilot's mental workload sensor technology. With some sketches of several possible designs, accompanied by brief descriptions of their functionalities, this rough design concept became the basis for FGD interview with pilots and the online experiment gathering information about the public's willingness-to-fly. Information gathered from these studies, particularly the FGD interview, could provide preliminary insights about design requirements. In human-computer interaction design, it is common to have a requirement activity to collect relevant data for developing the design concept (Sharp, 2019). Issues to be explored in requirements activity includes comprehending the target users and their capabilities, users' current tasks, goals, and contexts, or constraints on the device (Sharp, 2019). These issues seem to be relevant with information we have gathered from FGD interview in particular. We are therefore confident that our results would be beneficial in design processes of this technology in the future.

7.6 Limitations of the Research

The empirical results reported herein should be considered in the light of some limitations. We present limitations that are most likely to have impacts on the interpretation of our findings and conclusions. The limitations will be discussed according to the order of the study presented in this thesis.

Firstly, whilst the sample is numerically limited, we believe it is sufficient to generate useful insights and also note that recruiting highly skilled pilots is a challenging endeavour that generally precludes mass recruitment outside work initiated by airlines or aeroplane manufacturers themselves. The basic form of the TAM framework used in this study, based on our justification, also has advantages in providing a simple framework for a pioneering work in this issue. This consequently could reduce potential dropout from participants as the results of having too many variables. However, various extended versions of the TAM framework with more diverse variables might generate more insights. We also acknowledge some limitations regarding the online studies, such as technical errors during interviews and online experiment as well as biases in participant responses. Nevertheless, with some strategies applied to these studies, such as creating a list of questions for the interviews or altering participant's order during the FGD, we believe that we have minimised these potential errors and biases.

Secondly, whilst the CDM interview was an excellent tool to reveal the pilot's cognitive processes, some limitations must be acknowledged. The initial phase of the interview, which was the discussion regarding a sample event, was possibly the most crucial phase. This phase aims to agree the boundary of the event to be discussed later. Therefore, a general rule of starting and ending point of an event, such as the landing, had to be provided to minimise wide variations of the timeline created. In our study, this strategy, combined with reviewing the timeline and the interview excerpts by both the interviewer and the pilot, had generally worked in creating a more uniform timeline. However, some differences in the timeline still appeared due to interpretation of the pilots, particularly when determining the beginning of a landing phase. Whilst five pilots involved in this CDM interview were able to provide an ample amount of information, involving more pilots might result in more confidence in drawing conclusions. Still, disagreement among qualitative researcher regarding ideal number of participants in a qualitative study exists (Dworkin, 2012). Since this study relies on the framework from Sharples & Megaw (2015), we must also acknowledge the limitation pertinent to the framework. In particular, although the framework incorporates factors influencing MWL from the previous studies, this framework needs to be challenged further by scientific community for its robustness. Nevertheless, our study is one of the first studies that has attempted to investigate this framework empirically, thus, contributes to its development.

Thirdly, MATB is intended to generate the types of cognitive workload encountered while flying rather than to tap into expertise or domain knowledge held by pilots. Consequently, we believe the fundamental measurement of these workload types using fNIRS should generalise to a real flight environment, albeit that real pilots might have other specific strategies for managing their workload. Despite merely using general participants' data, our fNIRS studies had partially confirmed some previous studies regarding the use of fNIRS in measuring MWL during a flying task. In addition, despite the non-significant results, our study had possibly paved a way to understand the prospective MWL prediction during a multitasking environment.

With these identified limitations, the results of this research should be interpreted with caution as follows:

1. The attribution to the general population of professional pilots must be avoided. Pilots, to which the results of this research often referred, are limited to the group of pilots involved in the studies. Therefore, the results did not aim to represent all pilots and companies or organisations for which they work. This also applies for the context of members of the public.

2. The differences in mental workload changes as detected in the second experiment of the brain and physiological study (Chapter 5) require further research in order to apply these measures in operational contexts. Although it is possible to implement these devices in real-world flying situations, it needs further investigation to their validity and reliability. This also applies to the results regarding the vicarious observation of MWL (Chapter 6).

3. The hypothetical scenario developed for the acceptance study was generally based on the literature review regarding MWL measurement using brain and physiological sensors, with additional information assuming the technology would become operational in the future. The study therefore was inspired by general potential of the sensors yet not specifically attributed to results from specific studies.

7.7 Implications and Exploitation of the Research

The research in this thesis identifies several implications to consider. Discussion regarding implications of this research could benefit relevant parties in the sense of potential use of our research. We discuss these implications in three different perspectives, namely theoretical, methodological, and practical/domain-specific, as follows:

1. Regarding theoretical implications, our research supports the notion that subjective experience of MWL varies between individuals. Therefore, subjective experience of MWL during a task must be always considered, even when physiological measurements are the main methods utilised. From the results of our research, we also support the idea of MWL connection to certain physiological responses. Our results support the argument that MWL can be reflected in brain

activation, heart rate and RMSSD response, and pupil dilation. Despite various degree of the evidence, investigations of these physiological indicators of MWL have to be advanced so that the real implementation of it can be realised in the near future.

2. Concerning methodological implications, our research suggests that fNIRS is not sufficiently sensitive for detecting MWL in the context we explored. The results from our research can serve as the basis for further evaluation of the efficacy of fNIRS, and other physiological sensors, in detecting MWL in the aviation context. Our results also reveal benefits from the use of qualitative approaches to comprehending MWL in a flying task more deeply. The CDM interview can be used for identifying sources of task demands and their interaction with other elements of MWL, whilst acceptance studies using FGD in particular can reveal the nature of MWL as understood by professional pilots. The use of online methods is seemingly feasible for studying MWL within aviation context. Nevertheless, several methodological considerations have to be considered.

3. Regarding practical/domain-specific implications, our results may serve as the basis for providing training for MWL awareness among professional pilots. The reason behind this argument is that, from the pilots' perspective (as captured in Chapter 4), MWL is something that they tend to occasionally underestimate because of its subjectivity. MWL, in a real flight, can be interpreted independently by pilots, resulting in various strategies in decision-making or problem-solving. Furthermore, despite the insensitivity of fNIRS, introducing pilots to more objective methods of MWL detection in general can be initiated. For example, the curriculum for pilot training may include topics related to physiological measurements of MWL. As mentioned previously, since MWL tends to be interpreted subjectively, providing more objective metrics of MWL could help pilots to be more aware of their MWL status during a flight, potentially resulting in more standardised responses among pilots, thus improving operational safety.

7.8 Suggestions for Future Research

Based on the aforementioned limitations of this research, future research in similar areas should consider the following recommendations. Firstly, using a more ecologically valid flight simulator and inviting professional pilots might provide different insights. Mental workload changes measurement during a flying task, whether simulated in MATB or a more realistic flight simulator, could be more valid if involving professional pilots. This is because pilots have different strategies to predict and manage MWL of a flying task as the results of their experience and training. Further, the results would be more relevant and closer to solving the problem, and reach more valid conclusions regarding the efficacy of brain and other physiological sensors to detect MWL changes in a real-world context.

Secondly, procedures for experiments using brain and physiological measurements must be strategically designed, particularly if the study lasts for a longer period of time. Balancing between resting and experimental periods must be practically tested, for example, by conducting a pilot study with two or three participants with identical proposed procedures. Due to the idiosyncratic nature of brain and physiological measurements, including their preprocessing stages, this might be an iterative process.

Thirdly, time-series analysis of mental workload using qualitative techniques is worth consideration. If CDM interview techniques are to be used, strict timelines must be agreed before continuing the interview, since it will become the foundation for the analysis. More professional pilots to be invited may also result in stronger conclusions. We anecdotally suggest minimum 10 participants to be involved for a qualitative approach to increase chances of getting more valid results and to allow some operational errors. If one or two participants' data need to be dismissed, with 10 participants we would not lose so much data.

Fourthly, understanding pilots' attitudes towards MWL sensors technology is ideally undertaken in an environment that allows pilots to experience the technology or, at least, fully understand the concept of it. Creating a paper prototype and detailing the description of the scenarios could serve as the control for potential biases. More various type of aircraft that pilots fly might also enrich the information obtained from such study. In our case, we had one fighter jet pilot who, according to our impression, had provided quite unique perspectives of becoming a pilot and of responding to a new technology.

7.9 Concluding Statements

The research presented in this thesis has addressed several gaps within MWL literature, and reported findings which extend our understanding of MWL during a flying task by using various methods of measurements. Findings suggest that MWL during a flying task emerges from the dynamic interaction between task demands, performance, and external/internal influences throughout the flight. Furthermore, changes of MWL during this dynamic flying task can be partially detected by physiological indicators such as brain, heart rate, and pupil dilation. However, MWL of the task cannot be predicted before the task yet some subtasks constituting it could be, particularly, those with contrasting task elements such as system monitoring. With attempts to predict MWL more objectively during a real task have been proposed in previous studies, including this thesis for the context of aviation, the future implementation of MWL sensors technology appears to be likely. Findings from this thesis also revealed that both pilots and members of the public tend to agree with the idea of MWL sensors implementation in cockpit.

The thesis has therefore contributed critical knowledge to the understanding of what makes pilots experience high MWL during a flight and the way it was responded during the task. Practically speaking, this thesis has achieved an outcome regarding MWL measurement during a flying task by testing the efficacy of potential physiological sensors in ecologically valid environment. It is now essential that future research contribute to the extension of the context in measuring MWL during a flying task, with some considerations have to be considered as suggested by this thesis. By testing the sensors in more ecologically valid environment, such as in a flight simulator using real pilots, more valid conclusions regarding the efficacy of physiological sensors to detect MWL changes in real-world context can be reached; the development of sensors to be integrated in future cockpit can also be initiated. This endeavour, eventually, will generally improve safety in aviation.

References

Ahlstrom, U., Ohneiser, O., & Caddigan, E. (2016). Portable Weather Applications for General Aviation Pilots. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58(6), 864–885.

Ahn, S., Nguyen, T., Jang, H., Kim, J. G., & Jun, S. C. (2016). Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and fNIRS Data. *Frontiers in Human Neuroscience*, 10, 219.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.

Alaimo, A., Esposito, A., Orlando, C., & Simoncini, A. (2020). Aircraft Pilots Workload Analysis: Heart Rate Variability Objective Measures and NASA-Task Load Index Subjective Evaluation. *Aerospace*, 7(9), 137.

Allsop, J., Gray, R., Bulthoff, H. H., & Chuang, L. (2016). Effects of anxiety and cognitive load on instrument scanning behavior in a flight simulation. In *2016 IEEE Second Workshop on Eye Tracking and Visualization (ETVIS)* (pp. 55–59).

Anser, M. K., Yousaf, Z., & Zaman, K. (2020). Green Technology Acceptance Model and Green Logistics Operations: "To See Which Way the Wind Is Blowing". *Frontiers in Sustainability*, 1, 3.

Argyle, E. M., Marinescu, A., Wilson, M. L., Lawson, G., & Sharples, S. (2021). Physiological indicators of task demand, fatigue, and cognition in future digital manufacturing environments. *International Journal of Human-Computer Studies*, 145, 102522.

Ashraf, A. R., Thongpapanl, N. T., & Auh, S. (2014). The Application of the Technology Acceptance Model under Different Cultural Contexts: The Case of Online Shopping Adoption. *Journal of International Marketing*, 22(3), 68–93.

Asmayawati, S. & Nixon, J. (2020). Modelling and supporting flight crew decision-making during aircraft engine malfunctions: developing design recommendations from cognitive work analysis. *Applied Ergonomics*, 82, 102953.

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage*, 59(1), 36–47.

Battaglia, M. (2008). Purposive sample. In P. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 645–647). Thousand Oaks, CA: Sage Publications, Inc.

Behrendt, H. F., Firk, C., Nelson, C. A., & Perdue, K. L. (2018). Motion correction for infant functional near-infrared spectroscopy with an application to live interaction data. *Neurophotonics*, 5(1), 015004.

Bell, E. C., Willson, M. C., Wilman, A. H., Dave, S., & Silverstone, P. H. (2006). Males and females differ in brain activation during cognitive tasks. *NeuroImage*, 30(2), 529–538.

Bilimoria, K. D., Johnson, W. W., & Schutte, P. C. (2014). Conceptual framework for single pilot operations. In *Proceedings of the International Conference on Human-Computer Interaction in Aerospace* (pp. 1–8). Santa Clara, California: ACM Press.

Bliss, J. P. (1997). Alarm reaction patterns by pilots as a function of reaction modality. *International Journal of Aviation Psychology*, 7(1), 1–14.

Boele-Vos, M. J., Commandeur, J. J. F., & Twisk, D. A. M. (2017). Effect of physical effort on mental workload of cyclists in real traffic in relation to age and use of pedelecs. *Accident Analysis & Prevention*, 105, 84–94.

Borghini, G., Arico, P., Graziani, I., Salinari, S., Sun, Y., Taya, F., Bezerianos, A., Thakor, N. V., & Babiloni, F. (2016). Quantitative Assessment of the Training Improvement in a Motor-Cognitive Task by Using EEG, ECG and EOG Signals. *Brain Topography*, 29(1), 149–161.

Brannon, N. G., Koubek, R. J., & Voss, D. (2008). Mechanisms of knowledge degradation in a resource management task. *Theoretical Issues in Ergonomics Science*, 9(1), 25–44.

Brennan, S. (1992). *An experimental report on rating scale descriptior sets for the instantaneous self assessment (ISA) recorder.* Technical Report DRA Technical Memorandum (CAD5) 92017, DRA Maritime Command and Control Division, Portsmouth.

Budd, J., Miller, B. S., Manning, E. M., Lampos, V., Zhuang, M., Edelstein, M., Rees, G., Emery, V. C., Stevens, M. M., Keegan, N., Short, M. J., Pillay, D., Manley, E., Cox, I. J., Heymann, D., Johnson, A. M., & McKendry, R. A. (2020). Digital technologies in the public-health response to COVID-19. *Nature Medicine*, 26(8), 1183–1192.

Cain, B. (2007). *A Review of the Mental Workload Literature.* Technical report, Defence Research and Development Canada Toronto, Toronto.

Caldwell, J. A. & Ramspott, S. (1998). Effects of task duration on sensitivity to sleep deprivation using the multi-attribute task battery. *Behavior Research Methods Instruments & Computers*, 30(4), 651–660.

Carlozzi, N. E., Horner, M. D., Kose, S., Yamanaka, K., Mishory, A., Mu, Q., Nahas, Z., Wells, S. A., & George, M. S. (2010). Personality and Reaction Time after Sleep Deprivation. *Current Psychology*, 29(1), 24–33.

Cattermole, V. T., Horberry, T., & Hassall, M. (2016). Using Naturalistic Decision Making to Identify Support Requirements in the Traffic Incident Management Work Environment:. *Journal of Cognitive Engineering and Decision Making.*

Cattermole-Terzic, V. & Horberry, T. (2020). Improving Traffic Incident Management Using Team Cognitive Work Analysis. *Journal of Cognitive Engineering and Decision Making*, 14(2), 152–173.

Causse, M., Chua, Z., Peysakhovich, V., Campo, N. D., & Matton, N. (2017). Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS.

Scientific Reports, 7(1), 5222.

Cegarra, J., Valéry, B., Avril, E., Calmettes, C., & Navarro, J. (2020). OpenMATB: A Multi-Attribute Task Battery promoting task customization, software extensibility and experiment replicability. *Behavior Research Methods*, 52(5), 1980–1990.

Chau, P. Y. K. & Hu, P. J. (2002). Examining a Model of Information Technology Acceptance by Individual Professionals: An Exploratory Study. *Journal of Management Information Systems*, 18(4), 191–229.

Chiappe, D., Conger, M., Liao, J., Caldwell, J. L., & Vu, K.-P. L. (2013). Improving multi-tasking ability through action videogames. *Applied Ergonomics*, 44(2), 278–284.

Comstock, J. R. & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research.* Technical Report NASA-TM-104174, NAS 1.15:104174, NASA Langley Research Center, Hampton, VA.

Cooper, R. J., Selb, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., Ashina, M., & Boas, D. A. (2012). A Systematic Comparison of Motion Artifact Correction Techniques for Functional Near-Infrared Spectroscopy. *Frontiers in Neuroscience*, 6(147), 1–10.

Cowan, N. (2010). The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? *Current directions in psychological science*, 19(1), 51–57.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: a practitioner's guide to cognitive task analysis.* Cambridge, Mass: MIT Press.

Creswell, J. W. & Poth, C. N. (2018). *Qualitative inquiry & research design: choosing among five approaches.* London: SAGE Publications, 4th edition.

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage*, 54(4), 2808–2821.

Dahlstrom, N. & Nahlinder, S. (2009). Mental Workload in Aircraft and Simulator During Basic Civil Aviation Training. *The International Journal of Aviation Psychology*, 19(4), 309–325.

Dahlstrom, N., Nahlinder, S., Wilson, G. F., & Svensson, E. (2011). Recording of Psychophysiological Data During Aerobatic Training. *The International Journal of Aviation Psychology*, 21(2), 105–122.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319–340. Publisher: Management Information Systems Research Center, University of Minnesota.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35(8), 982–1003.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and Intrinsic Motivation to Use Computers in the Workplace1. *Journal of Applied Social Psychology*, 22(14), 1111–1132.

Davis, F. D. & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: three experiments. *International Journal* *of Human-Computer Studies*, 45(1), 19–45.

Delliaux, S., Delaforge, A., Deharo, J.-C., & Chaumet, G. (2019). Mental Workload Alters Heart Rate Variability, Lowering Non-linear Dynamics. *Frontiers in Physiology*, 10.

Di Lorenzo, R., Pirazzoli, L., Blasi, A., Bulgarelli, C., Hakuno, Y., Minagawa, Y., & Brigadoi, S. (2019). Recommendations for motion correction of infant fNIRS data applicable to multiple data sets and acquisition systems. *NeuroImage*, 200, 511–527.

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A Random Glance at the Flight Deck: Pilots' Scanning Strategies and the Real-Time Assessment of Mental Workload. *Journal of Cognitive Engineering and Decision Making*, 1(3), 271–285.

Diop, E. B., Zhao, S., & Duy, T. V. (2019). An extension of the technology acceptance model for understanding travelers' adoption of variable message signs. *PLOS ONE*, 14(4), e0216007.

Durantin, G., Scannella, S., Gateau, T., Delorme, A., & Dehais, F. (2016). Processing Functional Near Infrared Spectroscopy Signal with a Kalman Filter to Assess Working Memory during Simulated Flight. *Frontiers in Human Neuroscience*, 9(707), 1–9.

Dworkin, S. L. (2012). Sample Size Policy for Qualitative Studies Using In-Depth Interviews. *Archives of Sexual Behavior*, 41(6), 1319–1320.

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, 25, 69–91.

Edwards, T., Sharples, S., Wilson, J. R., & Kirwan, B. (2012). Factor interaction influences on human performance in air traffic control: The need for a multifactorial model. *Work*, 41, 159–166.

Endsley, M. R. & Bolstad, C. A. (1994). Individual Differences in Pilot Situation Awareness. *The International Journal of Aviation Psychology*, 4(3), 241–264.

Everly, J. J. (2016). Human Performance on a Signal Detection Task: Discriminability and Sensitivity to Reinforcement. *The Psychological Record*, 66(1), 139–151.

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, 56(2), 171–184.

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., Farhadian, M., & Miyake, S. (2016). Analysis of the mental workload of city traffic control operators while monitoring traffic density: A field study. *International Journal of Industrial Ergonomics*, 54, 170–177.

Fantini, S., Sassaroli, A., Tgavalekos, K. T., & Kornbluth, J. (2016). Cerebral blood flow and autoregulation: current measurement techniques and prospects for noninvasive optical methods. *Neurophotonics*, 3(3), 031411.

Ferraro, J., Christy, N., & Mouloua, M. (2017). Impact of Auditory Interference on Automated Task Monitoring and Workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 1136–1140.

Filimonau, V., Mika, M., & Pawlusiński, R. (2018). Public attitudes to biofuel use in aviation: Evidence from an emerging tourist market. *Journal of Cleaner Production*, 172, 3102–3110.

Foy, H. J. & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics*, 73, 90–99.

Foy, H. J., Runham, P., & Chapman, P. (2016). Prefrontal Cortex Activation and Young Driver Behaviour: A fNIRS Study. *PLOS ONE*, 11(5), 1–18.

Frederick-Recascino, C. M. & Hall, S. (2003). Pilot Motivation and Performance: Theoretical and Empirical Relationships. *The International Journal of Aviation Psychology*, 13(4), 401–414.

Frith, C. & Dolan, R. (1996). The role of the prefrontal cortex in higher cognitive functions. *Cognitive Brain Research*, 5(1), 175–181.

Fussell, S. G. & Truong, D. (2021). Using virtual reality for dynamic learning: an extended technology acceptance model. *Virtual Reality*, 26, 249–267.

Gagnon, L., Smith, A. F., Boas, D. A., Devor, A., Secomb, T. W., & Sakadžić, S. (2016). Modeling of Cerebral Oxygen Transport Based on In vivo Microscopic Imaging of Microvascular Network Structure, Blood Flow, and Oxygenation. *Frontiers in Computational Neuroscience*, 10(82), 1–20.

Galoyan, T., Betts, K., Abramian, H., Reddy, P., Izzetoglu, K., & Shewokis, P. A. (2021). Examining Mental Workload in a Spatial Navigation Transfer Game via Functional near Infrared Spectroscopy. *Brain Sciences*, 11(1), 45. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

Gibb, G. D. & Dolgin, D. L. (1989). Predicting Military Flight Training Success by a Compensatory Tracking Task. *Military Psychology*, 1(4), 235–240. Publisher: Routledge _eprint: https://doi.org/10.1207/s15327876mp0104_4.

Gilchrist, A. L., Cowan, N., & Naveh-Benjamin, M. (2008). Working Memory Capacity for Spoken Sentences Decreases with Adult Aging: Recall of Fewer, but not Smaller Chunks in Older Adults. *Memory*, 16(7), 773–787.

Glotzbach, E., Mühlberger, A., Gschwendtner, K., Fallgatter, A. J., Pauli, P., & Herrmann, M. J. (2011). Prefrontal Brain Activation During Emotional Processing: A Functional Near Infrared Spectroscopy Study (fNIRS). *The Open Neuroimaging Journal*, 5, 33–39.

Goodwin, L. D. & Leech, N. L. (2006). Understanding Correlation: Factors That Affect the Size of r. *The Journal of Experimental Education*, 74(3), 251–266.

Grös, D. F., Antony, M. M., Simms, L. J., & McCabe, R. E. (2007). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory (STAI). *Psychological Assessment*, 19(4), 369–381.

Gunzler, D., Chen, T., Wu, P., & Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. *Shanghai Archives of Psychiatry*, 25(6), 390–394.

Gutzwiller, R. S., Wickens, C. D., & Clegg, B. A. (2016). The Role of Time on Task in Multi-task Management. *Journal of Applied Research in Memory and Cognition*, 5(2), 176–184.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating Cognitive Capacity from

Knowledge: A New Hypothesis. *Trends in cognitive sciences*, 11(6), 236–242.

Hancock, G., Longo, L., Young, M., & Hancock, P. (2021). Mental Workload. In G. Salvendy & W. Karwowski (Eds.), *Handbook of Human Factors and Ergonomics* (pp. 203–226). Hoboken, NJ: John Wiley & Sons, Ltd, fifth edition.

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

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Oxford: North-Holland Press.

Hedge, J. W., Bruskiewicz, K. T., Borman, W. C., Hanson, M. A., Logan, K. K., & Siem, F. M. (2000). Selecting Pilots With Crew Resource Management Skills. *The International Journal of Aviation Psychology*, 10(4), 377–392.

Helleberg, J. R. & Wickens, C. D. (2003). Effects of Data-Link Modality and Display Redundancy on Pilot Performance: An Attentional Perspective. *The International Journal of Aviation Psychology*, 13(3), 189–210.

Herold, F., Wiegel, P., Scholkmann, F., & Müller, N. G. (2018). Applications of Functional Near-Infrared Spectroscopy (fNIRS) Neuroimaging in Exercise–Cognition Science: A Systematic, Methodology-Focused Review. *Journal of Clinical Medicine*, 7(12).

Hilgarter, K. & Granig, P. (2020). Public perception of autonomous vehicles: A qualitative study based on interviews after riding an autonomous shuttle. *Transportation Research Part F: Traffic Psychology and Behaviour*, 72, 226–243.

Hodgetts, H. M. & Jones, D. M. (2006). Interruption of the Tower of London task: Support for a goal-activation approach. *Journal of Experimental Psychology: General*, 135(1), 103–115.

Hollnagel, E. (2021). The Changing Nature of Task Analysis. In G. Salvendy & W. Karwowski (Eds.), *Handbook of Human Factors and Ergonomics* (pp. 358–367). Hoboken, NJ: John Wiley & Sons, Ltd, fifth edition.

Hooper, D., Coughlan, J., & Michael, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6(1).

Hsieh, H.-F. & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, 15(9), 1277–1288.

Hsu, B.-W., Wang, M.-J. J., Chen, C.-Y., & Chen, F. (2015). Effective Indices for Monitoring Mental Workload While Performing Multiple Tasks. *Perceptual and Motor Skills*, 121(1), 94–117.

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*, 48(10), 280–298.

Huppert, T. J., Hoge, R. D., Dale, A. M., Franceschini, M. A., & Boas, D. A. (2006a). Quantitative spatial comparison of diffuse optical imaging with blood oxygen level-dependent and arterial spin labeling-based functional magnetic resonance imaging. *Journal of Biomedical Optics*, 11(6), 064018.

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006b). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, 29(2), 368–382.

Jaquess, K. J., Lo, L.-C., Oh, H., Lu, C., Ginsberg, A., Tan, Y. Y., Lohse, K. R., Miller, M. W., Hatfield, B. D., & Gentili, R. J. (2018). Changes in Mental Workload and Motor Performance Throughout Multiple Practice Sessions Under Various Levels of Task Difficulty. *Neuroscience*, 393, 305–318.

Jöbsis, F. F. (1977). Noninvasive, Infrared Monitoring of Cerebral and Myocardial Oxygen Sufficiency and Circulatory Parameters. *Science*, 198(4323), 1264–1267.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall Inc.

Kahneman, D., Sibony, O., & Sunstein, C. (2022). *Noise: A Flaw in Human Judgment*. Glasgow: William Collins.

Kantenbacher, J., Hanna, P., Cohen, S., Miller, G., & Scarles, C. (2018). Public attitudes about climate policy options for aviation. *Environmental Science & Policy*, 81, 46–53.

Kantowitz, B. H. (2000). Attention and Mental Workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(21), 3–456–3–459.

Kelley, K., Clark, B., Brown, V., & Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in Health Care*, 15(3), 261–266.

Kennedy, L. & Parker, S. H. (2017). Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task. In *Proceedings of the 2017 International Symposium on Human Factors and Ergonomics in Health Care*, volume 6 (pp. 201–208). New Orleans, LA: Human Factors and Ergonomics Society.

Kharoufah, H., Murray, J., Baxter, G., & Wild, G. (2018). A review of human factors causations in commercial air transport accidents and incidents: From to 2000–2016. *Progress in Aerospace Sciences*, 99, 1–13.

Kim, J. H. & Yang, X. (2017). Applying fractal analysis to pupil dilation for measuring complexity in a process monitoring task. *Applied Ergonomics*, 65, 61–69.

Kim, J.-Y., Cho, Y.-J., & Min, S.-N. (2015). Psychophysiological measurement of physical and cognitive work. In J. R. Wilson & S. Sharples (Eds.), *Evaluation of Human Work Fourth Edition* (pp. 583–608). Boca Raton, FL: CRC Press, fourth edition.

Kim, Y. J., Chun, J. U., & Song, J. (2009). Investigating the role of attitude in technology acceptance from an attitude strength perspective. *International Journal of Information Management*, 29(1), 67–77.

Kirwan, B., Evans, A., Donohoe, L., Kilner, A., Lamoureux, T., Atkinson, T., & MacKendrick, H. (1997). Human Factors in the ATM System Design Life Cycle. In *Human Factors in the ATM System Design Life Cycle Report* Paris.

Kline, P. (1999). *Handbook of Psychological Testing*. London: Taylor & Francis Group.

Kojima, T., Tsunashima, H., Shiozawa, T., Takada, H., & Sakai, T. (2005). Measurement of Train Driver's Brain Activity by Functional Near-Infrared Spectroscopy (fNIRS). *Optical and Quantum Electronics*, 37(13), 1319–1338.

Kolb, B. & Whishaw, I. Q. (2016). Neuropsychology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology*, Cambridge Handbooks in

Psychology (pp. 26–38). Cambridge: Cambridge University Press, fourth edition.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4.

Lamb, T. L., Winter, S. R., Rice, S., Ruskin, K. J., & Vaughn, A. (2020). Factors that predict passengers willingness to fly during and after the COVID-19 pandemic. *Journal of Air Transport Management*, 89, 101897.

Lauren-Metz, K. (2019). Personal communication with the authors of "Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task".

Lei, S., Toriizuka, T., & Roetting, M. (2017). Driver adaptive task allocation: A field driving study. *Le Travail Humain*, 80(1), 93–112.

Lin, S.-Y., Juan, P.-J., & Lin, S.-W. (2020). A TAM Framework to Evaluate the Effect of Smartphone Application on Tourism Information Search Behavior of Foreign Independent Travelers. *Sustainability*, 12(22), 9366.

Linneberg, M. S. & Korsgaard, S. (2019). Coding qualitative data: a synthesis guiding the novice. *Qualitative Research Journal*, 19(3), 259–270.

Linton, P. M., Plamondon, B. D., Dick, A. O., Bittner, A. C., & Christ, R. E. (1989). Operator Workload for Military System Acquisition. In G. R. McMillan, D. Beevis, E. Salas, M. H. Strub, R. Sutton, & L. Van Breda (Eds.), *Applications of Human Performance Models to System Design* (pp. 21–45). Boston, MA: Springer US.

Lopez, N., Previc, F. H., Fischer, J., Heitz, R. P., & Engle, R. W. (2012). Effects of sleep deprivation on cognitive performance by United States Air Force pilots. *Journal of Applied Research in Memory and Cognition*, 1(1), 27–33.

Lu, C.-F., Liu, Y.-C., Yang, Y.-R., Wu, Y.-T., & Wang, R.-Y. (2015). Maintaining Gait Performance by Cortical Activation during Dual-Task Interference: A Functional Near-Infrared Spectroscopy Study. *PLOS ONE*, 10(6), e0129390.

MacSween-George, S. (2003). Will the public accept UAVs for cargo and passenger transportation? In *IEEE Aerospace Conference Proceedings*, volume 1 (pp. 1–367). Big Sky, MT: IEEE.

Maior, H. A., Pike, M., Sharples, S., & Wilson, M. L. (2015). Examining the Reliability of Using fNIRS in Realistic HCI Settings for Spatial and Verbal Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (pp. 3039–3042). Seoul, Republic of Korea: ACM Press.

Maior, H. A., Wilson, M. L., & Sharples, S. (2018). Workload Alerts—Using Physiological Measures of Mental Workload to Provide Feedback During Tasks. *ACM Trans. Comput.-Hum. Interact.*, 25(2), 9:1–9:30.

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3), 354–381.

Mansikka, H., Virtanen, K., Harris, D., & Simola, P. (2016). Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test. *Applied Ergonomics*, 56, 213–219.

Marinescu, A. C., Sharples, S., Ritchie, A. C., Sánchez López, T., McDowell, M., &

Morvan, H. P. (2018). Physiological Parameter Response to Variation of Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 60(1), 31–56.

Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of Eye-related Measures of Drivers' Mental Workload. *Procedia Manufacturing*, 3, 2854–2861.

Martínez-Mesa, J., González-Chica, D. A., Duquia, R. P., Bonamigo, R. R., & Bastos, J. L. (2016). Sampling: how to select participants in my research study? *Anais Brasileiros de Dermatologia*, 91(3), 326–330.

McKechnie, S., Winklhofer, H., & Ennew, C. (2006). Applying the technology acceptance model to the online retailing of financial services. *International Journal of Retail & Distribution Management*, 34(4/5), 388–410.

McMullin, C. (2021). Transcription and Qualitative Methods: Implications for Third Sector Research. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*.

Mehta, R., Winter, S. R., Rice, S., & Edwards, M. (2021). Are passengers willing to ride on autonomous cruise-ships? *Maritime Transport Research*, 2, 100014.

Menekse Dalveren, G. G., Cagiltay, N. E., Ozcelik, E., & Maras, H. (2018). Insights From Pupil Size to Mental Workload of Surgical Residents: Feasibility of an Educational Computer-Based Surgical Simulation Environment (ECE) Considering the Hand Condition. *Surgical Innovation*, 25(6), 616–624.

Metzger, U. & Parasuraman, R. (2005). Automation in Future Air Traffic Management: Effects of Decision Aid Reliability on Controller Performance and Mental Workload. *Human Factors*, 47(1), 35–49.

Midha, S., Maior, H. A., Wilson, M. L., & Sharples, S. (2021). Measuring Mental Workload Variations in Office Work Tasks using fNIRS. *International Journal of Human-Computer Studies*, 147, 102580.

Mier, W. & Mier, D. (2015). Advantages in functional imaging of the brain. *Frontiers in Human Neuroscience*, 9, 249.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.

Molavi, B. & Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological Measurement*, 33(2), 259–270.

Moray, N., Ed. (1979). *Mental Workload: Its Theory and Measurement*. Boston, MA: Springer US.

Moritz, S., Klein, J. P., Desler, T., Lill, H., Gallinat, J., & Schneider, B. C. (2017). Neurocognitive deficits in schizophrenia. Are we making mountains out of molehills? *Psychological Medicine*, 47(15), 2602–2612.

Naseer, N. & Hong, K.-S. (2015). fNIRS-based brain-computer interfaces: a review. *Frontiers in Human Neuroscience*, 9(3), 1–15.

Navon, D. & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86(3), 214–255.

Nelson, J., McKinley, R. A., Phillips, C., McIntire, L., Goodyear, C., Kreiner, A., & Monforton, L. (2016). The Effects of Transcranial Direct Current Stimulation (tDCS) on Multitasking Throughput Capacity. *Frontiers in Human Neuroscience*, 10(589), 1–13.

Nixon, J. & Charles, R. (2017). Understanding the human performance envelope using electrophysiological measures from wearable technology. *Cognition, Technology & Work*, 19(4), 655–666.

Norman, D. A. & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7(1), 44–64.

Pickup, L., Wilson, J., & Lowe, E. (2010). The Operational Demand Evaluation Checklist (ODEC) of workload for railway signalling. *Applied Ergonomics*, 41(3), 393–402.

Pickup, L., Wilson, J. R., Sharpies, S., Norris, B., Clarke, T., & Young, M. S. (2005). Fundamental examination of mental workload in the rail industry. *Theoretical Issues in Ergonomics Science*, 6(6), 463–482.

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., & Tachtsidis, I. (2019). Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Frontiers in Human Neuroscience*, 12(505), 1–21.

Podger, A. (2020). Public sector use of new technology: opportunities and challenges. *Asia Pacific Journal of Public Administration*, 42(4), 207–208.

Raby, M. & Wickens, C. D. (1994). Strategic Workload Management and Decision Biases in Aviation. *The International Journal of Aviation Psychology*, 4(3), 211–240.

Remington, R. & Loft, S. (2015). Attention and multitasking. In D. A. Boehm-Davis, F. T. Durso, & J. D. Lee (Eds.), *APA Handbook of Human System Integration* (pp. 261–276). American Psychological Association.

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, IL: Northwestern University.

Rice, S., Winter, S., Capps, J., Trombley, J., Robbins, J., Milner, M., & Lamb, T. (2020). Creation of Two Valid Scales: Willingness to Fly in an Aircraft and Willingness to Pilot an Aircraft. *International Journal of Aviation, Aeronautics, and Aerospace*, 7(1).

Rice, S., Winter, S. R., Mehta, R., & Ragbir, N. K. (2019). What factors predict the type of person who is willing to fly in an autonomous commercial airplane? *Journal of Air Transport Management*, 75, 131–138.

Richard, D., Buro, K., & Su, W. (2018). zeroEQpart-package: Zero Order vs (Semi) Partial Correlation Test and CI.

Richardson, C., Truong, D., & Choi, W. J. (2019). Examination of Factors Related to Pilot Acceptance Behaviors Toward the Automatic Ground Collision Avoidance System in Fighter Aircraft Operations. *The International Journal of Aerospace Psychology*, 29(1-2), 28–41.

Roberts, S. C., Ghazizadeh, M., & Lee, J. D. (2012). Warn me now or inform me later: Drivers' acceptance of real-time and post-drive distraction mitigation systems. *International Journal of Human-Computer Studies*, 70(12), 967–979.

Roese, N. J. & Vohs, K. D. (2012). Hindsight Bias. *Perspectives on Psychological Science*, 7(5), 411–426.

Rogers, E. (1995). *The Diffusion of Technology*. New York: The Free Press.

Rosseel, Y. (2012). lavaan: An *R* Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2).

Rowley, J. (2012). Conducting research interviews. *Management Research Review: MRN*, 35(3/4), 260–271.

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, 17(4), 688–690.

Sahaï, A., Barré, J., & Bueno, M. (2021). Urgent and non-urgent takeovers during conditional automated driving on public roads: The impact of different training programmes. *Transportation Research Part F: Traffic Psychology and Behaviour*, 81, 130–143.

Sato, H., Fuchino, Y., Kiguchi, M., Katura, T., Maki, A., Yoro, T., & Koizumi, H. (2005). Intersubject variability of near-infrared spectroscopy signals during sensorimotor cortex activation. *Journal of Biomedical Optics*, 10(4), 044001.

Sauvet, F., Jouanin, J. C., Langrume, C., Van Beers, P., Papelier, Y., & Dussault, C. (2009). Heart Rate Variability in Novice Pilots During and After a Multi-Leg Cross-Country Flight. *Aviation, Space, and Environmental Medicine*, 80(10), 862–869.

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, 128, 13–35.

Schmitz, F. & Voss, A. (2014). Components of task switching: A closer look at task switching and cue switching. *Acta Psychologica*, 151, 184–196.

Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1–66.

Schoedel, R., Hilbert, S., Bühner, M., & Stachl, C. (2018). One way to guide them all: Wayfinding strategies and the examination of gender-specific navigational instructions in a real-driving context. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58, 754–768.

Scholkmann, F. & Wolf, M. (2013). General equation for the differential pathlength factor of the frontal human head depending on wavelength and age. *Journal of Biomedical Optics*, 18(10), 105004.

Schutte, P. C. & Trujillo, A. C. (1996). Flight Crew Task Management in Non-Normal Situations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 40(4), 244–248.

Schöpfel, J. & Azeroual, O. (2021). 2 - Current research information systems and institutional repositories: From data ingestion to convergence and merger. In D. Baker & L. Ellis (Eds.), *Future Directions in Digital Information*, Chandos Digital Information Review (pp. 19–37). Chandos Publishing.

Shaffer, F. & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, 5(258), 1–17.

Sharp, H. (2019). *Interaction Design: Beyond Human-Computer Interaction*. Indianapolis, IN: John Wiley and Sons, fifth edition.

Sharples, S. (2019). Workload II: A Future Paradigm for Analysis and Measurement. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of* *the 20th Congress of the International Ergonomics Association (IEA 2018)*, Advances in Intelligent Systems and Computing (pp. 489–498). Cham: Springer International Publishing.

Sharples, S. & Cobb, S. (2015). Method for collecting and observing participant responses. In J. R. Wilson & S. Sharples (Eds.), *Evaluation of Human Work* (pp.$\tilde{8}$8). Boca Raton, FL: CRC Press, fourth edition.

Sharples, S. & Houghton, R. J. (2017). The field becomes the laboratory? The impact of the contextual digital footprint on the discipline of E/HF. *Ergonomics*, 60(2), 270–283.

Sharples, S. & Megaw, T. (2015). Definition and measurement of mental workload. In J. R. Wilson & S. Sharples (Eds.), *Evaluation of Human Work Fourth Edition* (pp. 515–548). Boca Raton, FL: CRC Press, fourth edition.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84(2), 127–190.

Smith, B. T. & Hess, T. M. (2015). The Impact of Motivation and Task Difficulty on Resource Engagement: Differential Influences on Cardiovascular Responses of Young and Older Adults. *Motivation science*, 1(1), 22–36.

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2013). *Human Factors Methods*. Surrey: Ashgate, second edition.

Stewart, N. & Harris, D. (2019). Passenger Attitudes to Flying on a Single-Pilot Commercial Aircraft. *Aviation Psychology and Applied Human Factors*, 9(2), 77–85.

Sublette, M., Carswell, C., Grant, R., Seidelman, W., Clark, D., & Seales, B. (2010). Anticipating Workload: Which Facets of Task Difficulty are Easiest to Predict? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54(19), 1704–1708.

Sublette, M., Carswell, C. M., Grant, R., Klein, M., Seales, W. B., & Clarke, D. (2009). Anticipated vs. Experienced Workload: How Accurately Can People Predict Task Demand? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Takae, Y., Yamamura, T., Kuge, N., Miyake, S., Yamada, S., & Hagiwara, C. (2010). Effects of Alcohol Intoxication on a Multi-task Simulator Operation Performance. *Review of Automotive Engineering*, 31(Final), 150–155.

Takeuchi, Y. (2000). Change in Blood Volume in the Brain during a Simulated Aircraft Landing Task. *Journal of Occupational Health*, 42(2), 60–65.

Taylor, S. & Todd, P. A. (1995). Understanding Information Technology Usage: A Test of Competing Models. *Information Systems Research*, 6(2), 144–176.

Thayer, J. F. & Lane, R. D. (2009). Claude Bernard and the heart–brain connection: Further elaboration of a model of neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, 33(2), 81–88.

Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal Computing: Toward a Conceptual Model of Utilization. *MIS Quarterly*, 15(1), 125–143.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: *R* Package for Causal Mediation Analysis. *Journal of Statistical Software*, 59(5).

UKRI (2021). *Future flight vision and roadmap*. Technical report, UKRI.

Van Acker, B. B., Parmentier, D. D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition, Technology & Work*, 20(3), 351–365.

Vance, S. M. & Malik, A. S. (2015). Analysis of factors that may be essential in the decision to fly on fully autonomous passenger airliners. *Journal of Advanced Transportation*, 49(7), 829–854.

Veltman, J. A. & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41(5), 656–669.

Venkatesh, V. & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, 39(2), 273–315.

Venkatesh, V. & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2), 186–204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425–478.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, 36(1), 157–178.

Verdière, K. J., Roy, R. N., & Dehais, F. (2018). Detecting Pilot's Engagement Using fNIRS Connectivity Features in an Automated vs. Manual Landing Scenario. *Frontiers in Human Neuroscience*, 12(6), 1–14.

Wanyan, X., Zhuang, D., Lin, Y., Xiao, X., & Song, J.-W. (2018). Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation. *International Journal of Industrial Ergonomics*, 64, 1–7.

Wassenaar, E. B. & Van den Brand, J. G. H. (2005). Reliability of Near-Infrared Spectroscopy in People With Dark Skin Pigmentation. *Journal of Clinical Monitoring and Computing*, 19(3), 195–199.

Waterson, P. (2011). World War II and other historical influences on the formation of the Ergonomics Research Society. *Ergonomics*, 54(12), 1111–1129.

Waxenbaum, J. A., Reddy, V., & Varacallo, M. (2021). Anatomy, Autonomic Nervous System. In *StatPearls*. Treasure Island, FL: StatPearls Publishing.

Welch, B. L. (1947). The Generalization of 'Student's' Problem when Several Different Population Variances are Involved. *Biometrika*, 34(1/2), 28–35.

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors*, 50(3), 449–455.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2012). *Engineering Psychology and Human Performance*. London: Taylor & Francis Group.

Williamson, T. & Spencer, N. (1989). Development and operation of the Traffic Alert and Collision Avoidance System (TCAS). *Proceedings of the IEEE*, 77(11), 1735–1744.

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology*, 12(1), 3–18.

Wilson, G. F., Caldwell, J. A., & Russell, C. A. (2007). Performance and Psychophysiological Measures of Fatigue Effects on Aviation Related Tasks of Varying Difficulty. *The International Journal of Aviation Psychology*, 17(2), 219–247.

Winter, S. R., Milner, M. N., Rice, S., Bush, D., Marte, D. A., Adkins, E., Roccasecca, A., Rosser, T. G., & Tamilselvan, G. (2018a). Pilot performance comparison between electronic and paper instrument approach charts. *Safety Science*, 103, 280–286.

Winter, S. R., Rice, S., Mehta, R., Walters, N. W., Pierce, M. B., Anania, E. C., Milner, M. N., & Rao, N. (2018b). Do Americans differ in their willingness to ride in a driverless bus? *Journal of Unmanned Vehicle Systems*, 6(4), 267–278.

Wolff, W., Schüler, J., Hofstetter, J., Baumann, L., Wolf, L., & Dettmers, C. (2019). Trait Self-Control Outperforms Trait Fatigue in Predicting MS Patients' Cortical and Perceptual Responses to an Exhaustive Task. *Neural Plasticity*, 2019, 1–10.

Wong, B. L. W. (2004). Critical Decision Method Data Analysis. In D. Diaper & N. A. Stanton (Eds.), *The Handbook of Task Analysis for Human-Computer Interaction* (pp. 327–346). Mahwah, NJ: Lawrence Erlbaum Associates.

Wright, K. B. (2005). Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *Journal of Computer-Mediated Communication*, 10(3), JCMC1034.

Wulvik, A. S., Dybvik, H., & Steinert, M. (2020). Investigating the relationship between mental state (workload and affect) and physiology in a control room setting (ship bridge simulator). *Cognition, Technology & Work*, 22(1), 95–108.

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

Young, M. S. & Stanton, N. A. (2002). Malleable Attentional Resources Theory: A New Explanation for the Effects of Mental Underload on Performance. *Human Factors*, 44(3), 365–375.

Zephyr Technology (2012). BioHarness 3.0 User Manual. Accessed from: https://www.zephyranywhere.com/media/download/bioharness3-user-manual.pdf.

Appendix A

Mathematical Expressions

A.1 MATB Performance Scoring (Study 5)

TRACK subtask score is defined as the root-mean-square deviation from the centre point in pixel units, and is mathematically expressed as:

$$TRACK = 1 - (\sqrt{(\frac{SS}{N})} \div 300) \tag{A.1}$$

where SS is the sum of the squares of the vertical (Y) and horizontal (X) offset of the target, with the maximum practical offset is 300 points. N is several samples during trial.

SYSMON subtask score was defined as the number of correct response from total stimuli introduced during the trial, and is mathematically expressed as:

$$SYSMON = \frac{\Sigma\, correct\ stimuli}{total\ stimuli} \tag{A.2}$$

RESMAN subtask score was defined as the sum of difference amount of fuel for both tanks from a designated tolerable deviation range (1000 units: 500 above and 500 below zero) during trial samples (recorded every 30 seconds), and is mathematically expressed as:

$$RESMAN = 1 - ((\frac{\Delta\, tank\ A}{1000}) + (\frac{\Delta\, tank\ B}{1000})) \tag{A.3}$$

Normalisation of the total score was performed using this formula:

$$x_{norm} = (\frac{x - x_{min}}{x_{max} - x_{min}}) \tag{A.4}$$

A.2 Brain Activation Index (Study 5)

Brain activation index was defined as the difference between oxygenated (HbOxy) and deoxygenated haemoglobin (HbDeoxy), and expressed mathematically as:

$$HbDiff = HbOxy - HbDeoxy \tag{A.5}$$

Appendix B

Design Concept and Questionnaires

B.1 Design Concept (Study 3)

The Future Technology - please read carefully

Research on mental workload (a psychological state indicating if you are too busy or not when doing your job) have advanced from capturing mind through simple subjective questionnaires to monitoring more objective physiological activities through measurement of brain, eye-gaze, or heart rate. The equipments to achieve the goal have also advanced, from bulky and noisy fMRI and lots-of-cables EEG device to more simple and portable device e.g. fNIRS, wearable heart rate monitoring tools, and eye-tracker.

Imagine following scenario happens to emerge in the future: Modern cockpit of every aircraft will integrate sensors that can monitor pilot's mental workload in real-time manner.

In terms of the equipment, if it is based on pilot's brain activity, it might be integrated with their headset so that there will be a small sensor touching small part of their forehead area, corresponding to prefrontal cortex area which is responsible for higher order cognitive activities such as thinking, predicting, etc. (see Image A).

Other possibilities would be in the form of a pair of glasses that can track their gazing behaviour, such as fixation, pupil dilatation, or blinks (Image B); or simply in the form of smartwatch that detects their heart rate measures, such as rates, variability, intervals, etc. (Image C).

In terms of the functionality, the sensors will detect pilot's physiological activities and, with advance data provision and robust algorithm (e.g. machine learning, deep learning, etc.), will give pilots real-time feedback about their current mental workload state. If it is crossing the (hypothetical) 'red line', meaning pilots are in the state of under- or overload, the sensor will alert and tell them what to do (e.g. "delegate some tasks").

A. Brain activity-based sensor B. Eye gaze behaviour-based sensor C. Heart rate-based sensor

B.2 Online Survey for Pilots (Study 3)

Opinion About Workload Sensors Technology in Cockpit

* Required

Introduction and Information Sheet

Aim of this study?

This study aims to gather opinion from professional pilots with regards to the implementation of future workload sensors technology in the cockpit.

How long and what about my rights and benefits?

The study will take no more than 30 minutes in total, in which you will be asked to complete a questionnaire. There will be no monetary compensation for your participation. Participation in this study is voluntary and you are free to stop the study at any time. If you choose to withdraw, any data collected will be stored with data from other participants.

What happen with my data?

All data collected will be analysed and stored in university's laptop and cloud service that is only accessible with login details by the researcher. The data will also be used for possible disseminations such as conferences, journal, and/or other forms of publication. In general, General Data Protection Regulation (GDPR) will be used for guidance in treating the data. More information about GDPR can be found here: https://www.nottingham.ac.uk/governance/records-and-information-management/gdpr-overview.aspx (https://www.nottingham.ac.uk/governance/records-and-information-management/gdpr-overview.aspx)

Who undertakes and funds this study?

This study is being completed as part of my roles as PhD student and Marie-Curie Early Stage Researcher co-funded by the European Union and The University of Nottingham.

Thank you for agreeing to participate. If you have any questions, don't hesitate to contact me (researcher). Contacts for my main supervisor and Faculty of Engineering Ethics Committee are also enclosed in case you need to reach them.

8/16/2021 Researcher

Ridwan Prasetyo
Marie Sklodowska-Curie Early Stage Researcher (ESR) & PhD Student
ezxrap@nottingham.ac.uk (mailto:ezxrap@nottingham.ac.uk)

Supervisor (Main)
Dr Robert Houghton
Associate Professor, Faculty of Engineering
robert.houghton@nottingham.ac.uk (mailto:robert.houghton@nottingham.ac.uk)

Faculty of Engineering Ethics Committee
ez-eng-ethics@nottingham.ac.uk (mailto:ez-eng-ethics@nottingham.ac.uk)

8/16/2021

Informed Consent

I have read the Participant Information Sheet and the nature and purpose of the research project has been explained to me. I understand and agree to take part.

I understand the purpose of the research project and my involvement in it.

I understand that my responses on this questionnaire are solely my own and not representing company, institution, or organisation I am currently working for.

I understand that I may withdraw from the research project at any stage and that this will not affect my status now or in the future.

I understand that while information gained during the study may be published, I will not be identified and my personal results will remain confidential.

I understand that data will be stored securely, treated in accordance with the General Data Protection Regulation (GDPR), and remain accessible only for researcher and his supervisors.

I understand that I may contact the researcher or supervisors if I require further information about the research, and that I may contact the Ethics Committee of the Faculty of Engineering, University of Nottingham, if I wish to make a complaint relating to my involvement in the research.

1

I understand the terms & conditions mentioned above and indicate my willingness to voluntarily take part in the study. *

☐ Tick to indicate your consent.

8/16/2021

Scenario

Research on mental workload (a psychological state indicating if you are too busy or not when doing your job) have advanced from capturing mind through simple subjective questionnaires to monitoring more objective physiological activities through measurement of brain, eye-gaze, or heart rate. The equipments to achieve the goal have also advanced, from bulky and noisy fMRI and lots-of-cables EEG device to more simple and portable device e.g. fNIRS, wearable heart rate monitoring tools, and eye-tracker.

Imagine following scenario happens to emerge in the future: Your modern cockpit will integrate sensors that can monitor your mental workload in real-time manner.

In terms of the equipment, if it is based on your brain activity, it might be integrated with your headset so that there will be a small sensor touching your small part of your forehead area (see Image A). Other possibilities would be in the form of a pair of glasses that can track your gazing behaviour (Image B), or simply in the form of smartwatch that detects your heart rate measures (Image C).

In terms of the functionality, the sensors will detect your physiological activities and, with advance data provision and robust algorithm (e.g. machine learning, deep learning, etc.), will give you real-time feedback about your current mental workload state. If it is crossing the (hypothetical) red line, meaning you are in the state of under- or overload, the sensor will alert you and tell you what to do (e.g. "delegate some tasks"). This system would act similar to what GPWS or TCAS would do to you.

Based on this scenario, and in addition to on-going and future improvements in cockpit technology such as touchscreens, head-up displays, and augmented reality, what do you think about this technology? Please proceed to the next page.

A. Brain activity-based sensor   B. Eye gaze behaviour-based sensor   C. Heart rate-based sensor

8/16/2021

Questions

Please rate each statement below. 1 indicating "strongly DISAGREE", meanwhile 5 indicating "strongly AGREE".

2

[Q1] Workload sensors technology will have positive impact. *

1 2 3 4 5

3

[Q2] I think it is a good idea to use sensors to monitor my workload. *

1 2 3 4 5

4

[Q3] Using workload sensors would be comfortable. *

1 2 3 4 5

5

[Q4] Workload sensors technology is beneficial for a flight operation. *

1 2 3 4 5

6

[Q5] I intend to use workload sensors in every flight I am commanding. *

1 2 3 4 5

7

[Q6] Using workload sensors would make me aware how busy I and my partner pilot are. *

1 2 3 4 5

8

[Q7] Using workload sensors would be easy. *

1 2 3 4 5

9

[Q8] I find it interesting to use sensors to monitor my workload. *

1 2 3 4 5

10

[Q9] I intend to recommend my flying partner to use workload sensors. *

1 2 3 4 5

11

[Q10] Using workload sensors would not obstruct my flying tasks. *

1 2 3 4 5

12

[Q11] Workload sensor would be reliable e.g. consistent in detecting workload dynamics. *

1 2 3 4 5

13

[Q12] Learning workload sensors attributes (how to use, what the feedback means, what to do after the feedback, etc.) would be easy. *

1 2 3 4 5

14

[Q13] I intend to use workload sensors once it is available in the cockpit. *

1 2 3 4 5

15

[Q14] Using workload sensors would make my flight safer. *

1 2 3 4 5

16

[Q15] I would trust what workload sensors alert me about my current workload. *

1 2 3 4 5

17

[Q16] Workload sensors would give clear and understandable feedback. *

1 2 3 4 5

Demographics

Please complete short questions below.

18

Date of completion *

Format: M/d/yyyy

19

Your total flight hours (all ratings) - please put approximate numbers. *

20

Your current rank - please choose the closest one if the choices below don't suit yours. *

- ◯ Captain (incl. Instructor)
- ◯ Senior First Officer
- ◯ First Officer
- ◯ Second Officer
- ◯ Trainee/Student/Cadet/Ab Initio

21

Your gender *

- ◯ Female
- ◯ Male

Invitation for my next study

I am going to undertake another study to deepen these responses in the form of focused group discussion (FGD). I would be very happy if you can join the study. There will be maximum 3 professional pilots at the same time and myself as facilitator per session. The discussion will last 90 minutes and be conducted online using Microsoft Teams.

Information sheet for the study can be found here: https://uniofnottm-my.sharepoint.com/:b:/g/personal/ridwan_prasetyo_nottingham_ac_uk/EYhMbc_Sy3BHrbhf3IWxZHoBirZFuGzssjiij0V906VpJA?e=YzAEtX (https://uniofnottm-my.sharepoint.com/:b:/g/personal/ridwan_prasetyo_nottingham_ac_uk/EYhMbc_Sy3BHrbhf3IWxZHoBirZFuGzssjiij0V906VpJA?e=YzAEtX)

In case you need to ask anything or discuss further about the study, you can drop me an email (ezxrap@nottingham.ac.uk (mailto:ezxrap@nottingham.ac.uk)).

If you are interested to join, please put your email address below and I will directly contact you to explain the study plan (you may say 'yes' now and withdraw this invitation later if you find the study is not suitable for you).

22

Your email address

Closing

You have completed the study and we thank you for your participation!

B.3 Online Survey for Public (Study 3)

Public Opinion About Workload Sensors Technology in Cockpit [S5QP1]

* Required

Introduction and Information Sheet

Aim of this study?

This study aims to gather opinion from general public with regards to the implementation of future workload sensors technology in the cockpit.

How long and what about my rights and benefits?

The study will take no more than 15 minutes in total, in which you will be presented with a scenario and complete a questionnaire. There will be no direct monetary compensation for your participation, but you can decide to join prize drawing of £10 Amazon voucher for 10 winners. Participation in this study is voluntary and you are free to stop the study at any time. If you choose to withdraw, any data collected will be stored with data from other participants.

What happen with my data?

All data collected will be analysed and stored in university's laptop and cloud service that is only accessible with login details by the researcher. The data will also be used for possible disseminations such as conferences, journal, and/or other forms of publication. In general, General Data Protection Regulation (GDPR) will be used for guidance in treating the data. More information about GDPR can be found here: https://www.nottingham.ac.uk/governance/records-and-information-management/gdpr-overview.aspx (https://www.nottingham.ac.uk/governance/records-and-information-management/gdpr-overview.aspx)

Who undertakes and funds this study?

This study is being completed as part of my roles as PhD student and Marie-Curie Early Stage Researcher co-funded by the European Union and The University of Nottingham.

Thank you for agreeing to participate. If you have any questions, don't hesitate to contact me (researcher). Contacts for my main supervisor and Faculty of Engineering Ethics Committee are also enclosed in case you need to reach them.

8/16/2021

Researcher
Ridwan Prasetyo
Marie Sklodowska-Curie Early Stage Researcher (ESR) & PhD Student
ezxrap@nottingham.ac.uk (mailto:ezxrap@nottingham.ac.uk)

Supervisor (Main)
Dr Robert Houghton
Associate Professor, Faculty of Engineering
robert.houghton@nottingham.ac.uk (mailto:robert.houghton@nottingham.ac.uk)

Faculty of Engineering Ethics Committee
ez-eng-ethics@nottingham.ac.uk (mailto:ez-eng-ethics@nottingham.ac.uk)

8/16/2021

Informed Consent

I have read the Participant Information Sheet and the nature and purpose of the research project has been explained to me. I understand and agree to take part.

I understand the purpose of the research project and my involvement in it.

I understand that my responses on this questionnaire are solely my own and not representing company, institution, or organisation I am currently working for.

I understand that I may withdraw from the research project at any stage and that this will not affect my status now or in the future.

I understand that while information gained during the study may be published, I will not be identified and my personal results will remain confidential.

I understand that data will be stored securely, treated in accordance with the General Data Protection Regulation (GDPR), and remain accessible only for researcher and his supervisors.

I understand that I may contact the researcher or supervisors if I require further information about the research, and that I may contact the Ethics Committee of the Faculty of Engineering, University of Nottingham, if I wish to make a complaint relating to my involvement in the research.

1

I understand the terms & conditions mentioned above and indicate my willingness to voluntarily take part in the study. *

☐ Tick to indicate your consent.

8/16/2021

Instruction

You will be presented with a scenario and then will be asked some questions about that scenario. Following that, you will be asked some demographics questions.

8/16/2021

The Future Technology - please read carefully

Research on mental workload (a psychological state indicating if you are too busy or not when doing your job) have advanced from capturing mind through simple subjective questionnaires to monitoring more objective physiological activities through measurement of brain, eye-gaze, or heart rate. The equipments to achieve the goal have also advanced, from bulky and noisy fMRI and lots-of-cables EEG device to more simple and portable device e.g. fNIRS, wearable heart rate monitoring tools, and eye-tracker.

Imagine following scenario happens to emerge in the future: Modern cockpit of every aircraft will integrate sensors that can monitor pilot's mental workload in real-time manner.

In terms of the equipment, if it is based on pilot's brain activity, it might be integrated with their headset so that there will be a small sensor touching small part of their forehead area, corresponding to prefrontal cortex area which is responsible for higher order cognitive activities such as thinking, predicting, etc. (see Image A).

Other possibilities would be in the form of a pair of glasses that can track their gazing behaviour, such as fixation, pupil dilatation, or blinks (Image B); or simply in the form of smartwatch that detects their heart rate measures, such as rates, variability, intervals, etc. (Image C).

In terms of the functionality, the sensors will detect pilot's physiological activities and, with advance data provision and robust algorithm (e.g. machine learning, deep learning, etc.), will give pilots real-time feedback about their current mental workload state. If it is crossing the (hypothetical) 'red line', meaning pilots are in the state of under- or overload, the sensor will alert and tell them what to do (e.g. "delegate some tasks").

A. Brain activity-based sensor B. Eye gaze behaviour-based sensor C. Heart rate-based sensor

Scenario

Now imagine that you are a passenger on a 2-hour commercial domestic flight. You are told that the pilots on board WILL BE using workload sensors technology to assist them in monitoring their workload.

2

Based on the scenario above, how strongly do you disagree or agree with these following statements? *

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
I would be willing to fly in this situation.	○	○	○	○	○
I would be comfortable flying in this situation.	○	○	○	○	○
I would have no problem flying in this situation.	○	○	○	○	○
I would be happy to fly in this situation.	○	○	○	○	○
I would feel safe flying in this situation.	○	○	○	○	○
I have no fear of flying in this situation.	○	○	○	○	○
I feel confident flying in this situation.	○	○	○	○	○

Scenario

Now imagine that you are a passenger on a 2-hour commercial domestic flight. You are told that the pilots on board WILL NOT BE using workload sensors technology to assist them in monitoring their workload.

2

Based on the scenario above, how strongly do you disagree or agree with these following statements? *

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
I would be willing to fly in this situation.	○	○	○	○	○
I would be comfortable flying in this situation.	○	○	○	○	○
I would have no problem flying in this situation.	○	○	○	○	○
I would be happy to fly in this situation.	○	○	○	○	○
I would feel safe flying in this situation.	○	○	○	○	○
I have no fear of flying in this situation.	○	○	○	○	○
I feel confident flying in this situation.	○	○	○	○	○

Demographics

Please answer these questions.

3

Your gender *

○ Female

○ Male

4

Your age *

5

How many flight do you usually take in a year? (Assume normal/non-pandemic situation) *

○ None to 4

○ 5 to 12

○ More than 12

Prize Drawing

If you wish, you can leave your email address below and get a chance to win £10 Amazon (UK) voucher for 10 winners.

However, as participants of this online study may be currently living outside the UK, the voucher may be inapplicable. In this case, if you win the prize, we can talk later to find a solution e.g. exchanging with local currency, etc.

Winners will be contacted directly via email; and email addresses generated from this section will be deleted immediately after the drawing.

If you don't want to join, just ignore this section.

6

Your email address

Closing

You have completed the study and we thank you for your participation!

B.4 TLX, ISA, MIAMI, and STICSA Scales (Study 5)

Level	Workload Heading	Spare Capacity	Description
1	Under-utilised	Very much	Nothing to do. Rather boring.
2	Relaxed	Ample	More than enough time for all tasks. Active on the task less than 50% of the time.
3	Comfortable Busy Pace	Some	All tasks well in hand. Busy but stimulating pace. Could keep going continuously at this level.
4	High	Very Little	Non-essential tasks suffering. Could not work at this level very long.
5	Excessive	None	Behind on tasks; losing track of the full picture.

Rating Scale Definitions

Title	Descriptions
MENTAL DEMAND	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Place a mark at the desired point on each scale:

MENTAL DEMAND
Low High

PHYSICAL DEMAND
Low High

TEMPORAL DEMAND
Low High

PERFORMANCE
Good Poor

EFFORT
Low High

FRUSTRATION
Low High

pre
Date:____________; Code:__________________

Questionnaire about neuropsychological assessment (MIAMI-R)

In the following you will find several questions which relate to the subsequent neuropsychological assessment as well as your performance. Please read each statement carefully and decide how far/ to which extent this statement applies to you. Indicate your answer for each statement by crossing out the number corresponding to your choice.

	fully agree	rather agree	rather disagree	fully disagree
1. I only take part in the neuropsychological assessment, because someone else (e.g., therapist, partner) insisted.	1	2	3	4
2. I worry that the tasks will be too difficult for me.	1	2	3	4
3. I know what I can expect from the assessment.	1	2	3	4
4. I am aware of my capabilities and will be able to show them in the test situation.	1	2	3	4
5. I can hardly concentrate today.	1	2	3	4
6. I feel fit and capable.	1	2	3	4
7. Test situations like this are not for me.	1	2	3	4
8. I don't take psychological tests too seriously.	1	2	3	4
9. I experience the test situation as very unpleasant and would like to leave.	1	2	3	4
10. I fear that a bad test result will have negative consequences.	1	2	3	4
11. I am so nervous that my performance will not reflect my true capabilities.	1	2	3	4
12. I experience sensations/complaints which are particularly severe and distracting today	1	2	3	4
13. I don't care much about the assessment.	1	2	3	4
14. I fear that I will perform poorly.	1	2	3	4
15. I am not motivated at all to take part in the assessment.	1	2	3	4
16. Right now, I feel very tired and exhausted.	1	2	3	4
17. It is important to me to perform well today.	1	2	3	4
18. I am willing to do my best.	1	2	3	4
19. I feel comfortable and in good hands with the tester.	1	2	3	4
20. I don't worry about the outcome of this assessment	1	2	3	4

Thank you!

pre
Date:____________; Code:__________________

Subscales
R= Items Reversed

1. poor motivation

1	I only take part in the neuropsychological assessment, because someone else (e.g., therapist, partner) insisted.	
13	I don't care much about the assessment.	R
15	I am not motivated at all to take part in the assessment.	
17	It is important to me to perform well today.	R
18	I am willing to do my best.	R

2. concerns about assessment

3	I know what I can expect from the assessment.	R
7	Test situations like this are really not for me.	
8	I don't take psychological tests too seriously.	
9	I experience the test situation as very unpleasant and would like to leave	
19NEW	I feel comfortable and in good hands with the tester.	R

3. fear about poor outcome

2	I worry that the tasks will be too difficult for me	
4	I am aware of my capabilities and will be able to show them in the test situation.	R
10	I fear that a bad test result will have negative consequences.	
14	I fear that I will perform poorly	
20NEW	I don't worry about the outcome of this assessment	R

4. negative momentary influences

5	I can hardly concentrate today.	
6	I feel fit and capable.	R
11	I am so nervous that my performance will not reflect my true capabilities.	
12	I experience sensations/complaints which are particularly severe and distracting today.	
16	Right now, I feel very tired and exhausted.	

post
Date:____________; Code:__________________

Questionnaire about neuropsychological assessment (MIAMI-R)

In the following you will find several questions which relate to the previous neuropsychological assessment as well as your performance. Please read each statement carefully and decide, how far/ to which extent this statement applies to you. Indicate your answer for each statement by crossing out the number corresponding to your choice.

	fully agree	rather agree	rather disagree	fully disagree
1. I pushed myself and gave my best.	1	2	3	4
2. I felt under pressure by the test situation.	1	2	3	4
3. The test situation was not as bad.	1	2	3	4
4. I felt very exhausted and worn out.	1	2	3	4
5. Many of the tasks were fun to do.	1	2	3	4
6. I was very nervous throughout the whole assessment.	1	2	3	4
7. I found many of the assigned tasks very difficult.	1	2	3	4
8. I could sufficiently concentrate on the tasks.	1	2	3	4
9. The assessment was way too long and in the end I lost my motivation.	1	2	3	4
10. I did not worry about the outcome of the assessment	1	2	3	4
11. There were too many things (e.g., noise) which disturbed my concentration.	1	2	3	4
12. During the assessment I repeatedly thought about how my performance would turn out.	1	2	3	4
13. I was not motivated and accordingly did not achieve my best performance.	1	2	3	4
14. I was fearful about the results while performing the tasks.	1	2	3	4
15. I felt being observed and uncomfortable, so I could not fully concentrate.	1	2	3	4
16. Different thoughts and concerns bothered and distracted me during the assessment.	1	2	3	4
17. I felt comfortable and in good hands with the tester.	1	2	3	4
18. I was distracted by bodily sensations (e.g., pain) during the assessment.	1	2	3	4
19. I did not really try hard	1	2	3	4
20. I was so focused on the tasks, I almost forgot that I was in a test situation	1	2	3	4

Thank you !

post

Date:_____________; Code:___________________

Subscales

R= Items Reversed

1. poor motivation

1	I pushed myself and gave my best.	R
5	Many of the tasks were fun to do.	R
9	The assessment was way too long and in the end I lost my motivation.	
13	I was not motivated and accordingly did not achieve my best performance.	
19NEW	I did not really try hard.	

2. concerns about assessment

2	I felt under pressure by the test situation.	
3	The test situation was not as bad.	R
11	There were too many things (e.g., noise) which disturbed my concentration	
15	I felt being observed and uncomfortable, so I could not fully concentrate.	
17New	I felt comfortable and in good hands with the tester.	R

3. fear about poor outcome

7	I found many of the assigned tasks very difficult.	
12	During the assessment I repeatedly thought about how my performance would turn out.	
14	I was fearful about the results while performing the tasks.	
10NEW	I did not worry about the outcome of the assessment.	R
20NEW	I was so focused on the tasks, I almost forgot that I was in a test situation.	R

4. negative momentary influences

4	I felt very exhausted and worn out.	
6	I was very nervous throughout the whole assessment.	
8	I could sufficiently concentrate on the tasks.	R
16	Different thoughts and concerns bothered and distracted me during the assessment.	
18	I was distracted by bodily sensations (e.g., pain) during the assessment.	

Participant ID: ______________________________

Your Mood at This Moment

Instructions

Below is a list of statements which can be used to describe how people feel. Beside each statement are four numbers which indicate the degree with which each statement is self-descriptive of mood at this moment (e.g., 1 = *not at all,* 4 = *very much so*). Please read each statement carefully and circle the number which best indicates how you feel right now, at this very moment, even if this is not how you usually feel.

	Not at All	A Little	Moderately	Very Much So
1. My heart beats fast.	1	2	3	4
2. My muscles are tense.	1	2	3	4
3. I feel agonized over my problems.	1	2	3	4
4. I think that others won't approve of me.	1	2	3	4
5. I feel like I'm missing out on things because I can't make up my mind soon enough.	1	2	3	4
6. I feel dizzy.	1	2	3	4
7. My muscles feel weak.	1	2	3	4
8. I feel trembly and shaky	1	2	3	4
9. I picture some future misfortune.	1	2	3	4
10. I can't get some thought out of my mind.	1	2	3	4
11. I have trouble remembering things.	1	2	3	4
12. My face feels hot.	1	2	3	4
13. I think that the worst will happen.	1	2	3	4
14. My arms and legs feel stiff.	1	2	3	4
15. My throat feels dry.	1	2	3	4
16. I keep busy to avoid uncomfortable thoughts.	1	2	3	4
17. I cannot concentrate without irrelevant thoughts intruding.	1	2	3	4
18. My breathing is fast and shallow.	1	2	3	4
19. I worry that I cannot control my thoughts as well as I would like to.	1	2	3	4
20. I have butterflies in the stomach.	1	2	3	4
21. My palms feel clammy.	1	2	3	4

Participant ID: ______________________________

Your General Mood State

Instructions

Below is a list of statements which can be used to describe how people feel. Beside each statement are four numbers which indicate how often each statement is true of you (e.g., 1 = *not at all,* 4 = *very much so*). Please read each statement carefully and circle the number which best indicates how you often, in general, the statement is true of you.

	Not at All	A Little	Moderately	Very Much So
1. My heart beats fast.	1	2	3	4
2. My muscles are tense.	1	2	3	4
3. I feel agonized over my problems.	1	2	3	4
4. I think that others won't approve of me.	1	2	3	4
5. I feel like I'm missing out on things because I can't make up my mind soon enough.	1	2	3	4
6. I feel dizzy.	1	2	3	4
7. My muscles feel weak.	1	2	3	4
8. I feel trembly and shaky	1	2	3	4
9. I picture some future misfortune.	1	2	3	4
10. I can't get some thought out of my mind.	1	2	3	4
11. I have trouble remembering things.	1	2	3	4
12. My face feels hot.	1	2	3	4
13. I think that the worst will happen.	1	2	3	4
14. My arms and legs feel stiff.	1	2	3	4
15. My throat feels dry.	1	2	3	4
16. I keep busy to avoid uncomfortable thoughts.	1	2	3	4
17. I cannot concentrate without irrelevant thoughts intruding.	1	2	3	4
18. My breathing is fast and shallow.	1	2	3	4
19. I worry that I cannot control my thoughts as well as I would like to.	1	2	3	4
20. I have butterflies in the stomach.	1	2	3	4
21. My palms feel clammy.	1	2	3	4