Development and evaluation of a behavioural marker system for UK fire and rescue service incident commanders

Philip Carl Butler

A thesis submitted to Cardiff University

For the Degree of Doctor of Philosophy

July 2021

Summary

This thesis developed and evaluated a behavioural marker system for UK Fire and Rescue Service (FRS) incident commanders (IC) to enable their non-technical skills to be assessed. Such skills complement technical skills and are essential to safety. Qualitative and quantitative methods were used to develop and evaluate the behavioural marker system (called THINCS). Evaluations found THINCS performed well with respect to its validity, reliability, sensitivity and usability. THINCS was then used together with other research methods to investigate how non-technical skills are deployed in different types of incident. The study involved simulated scenarios that required ICs to use standard operating procedures, or operational discretion. The results showed that these two approaches to decision making were not deployed in the ways anticipated by National Operational Guidance (N.O.G., 2021a). Instead, operational discretion was less likely to be used in the scenario in which it was licensed (Discretion scenario) than in the scenario in which it was not licensed (Control scenario). Also, ICs were more likely to exhibit acute stress in the Discretion scenario than in the Control scenario, as measured by self-report and a blood-based measure of immunological function. To help understand the decisions made, THINCS was used to examine differences in the use of non-technical skills in the two scenarios. The principal difference was in situational awareness, which was less evident in the Discretion than the Control scenario. Thematic analysis of post-scenario interviews showed that the ICs who used operational discretion in a context-appropriate manner demonstrated suitable judgement of risk and safety leadership compared to those who used it inappropriately. This thesis advocates that personal resilience training should focus on acute stress, and ICs

should be better prepared to use operational discretion. Finally, THINCS now features in the National Operational Guidance for UK incident command and is used by several UK FRSs.

Declaration

Statement 1: This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD

Signed: Date: 30/07/2021

Statement 2: This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is it being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation)

Signed: Date: 30/07/2021

Statement 3: I hereby give consent for my thesis, if accepted, to be available in the University's Open Access repository (or, where approved, to be available in the University's library and for inter-library loan), and for the title and summary to be made available to outside organisations, subject to the expiry of a University-approved bar on access if applicable.

Signed: Date: 30/07/2021

Declaration: This thesis is the result of my own independent work, except where otherwise stated, and the views expressed are my own. Other sources are acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third Party Editors by Research Degree Students Procedure.

Signed: Date: 30/07/2021

Dedication

I would like to dedicate this thesis to the 33 firefighters and other fire and rescue service personnel who tragically lost their lives responding to incidents during my 31 year career. Sadly, 8 of those who died were colleagues when I was serving in the Dorset and London Fire Brigades and I remember them here:

- Graham White (Dorset Fire Brigade) 1987
- Dave Gregory (Dorset Fire Brigade) 1987
- Terry Hunt (London Fire Brigade) 1991
- David Stokoe (London Fire Brigade) 1991
- Kevin Power (London Fire Brigade) 1992
- Michael Hill (London Fire Brigade) 1993
- Billy Faust (London Fire Brigade) 2004
- Adam Meere (London Fire Brigade) 2004

Acknowledgements

I would like to acknowledge the inspirational women and men who work as incident commanders for the UK fire and rescue services. To be *the* incident commander who takes charge of an emergency places huge responsibilities upon them to ensure the safety of firefighters, other emergency service personnel, and members of the public. With massive expectations that they will succeed in meeting those responsibilities. Consequently, the personal challenge laid down by these responsibilities and expectations is not for everyone. It takes great courage to step up and meet them head on every time, whilst figuring out how to resolve the incident. It is good to know that we have courageous people leading our brave firefighters.

A special recognition goes out to the 15 fire and rescue services that allowed me to 'borrow' their personnel to conduct my research in Hampshire. In particular, I would like to thank the East Sussex, Hampshire, Scottish, South Wales, South Yorkshire, West Yorkshire and Tyne and Wear FRSs. I shall be eternally grateful for the support of the magnificent former Hampshire FRS Deputy Chief Officers Andy Bowers and Steve Apter, the redoubtable former Oxfordshire FRS and currently London Fire Brigade Group Manager Jim Davies, and Avon FRS for the staunch Station Manager Darren Kingston. I cannot thank you all enough for your support, dedication, good senses of humour, and outstanding contributions along the way.

The support and encouragement I received from the National Fire Chiefs Council was outstanding. I could not have wished for better. It was a privilege to work with the formidable former South Yorkshire FRS Chief Fire Officer Jamie Courtney (QFSM), the forward thinking former Tyne and Wear FRS Assistant Chief Officer, John Baines, and the inspirational West Sussex FRS Chief Fire Officer, Dr Sabrina Cohen-Hatton (who also somehow found time to be part of my supervisory team). Likewise I thank the members of the National Operational Effectiveness Working Group and National Command and Control Users Group, whose generosity of spirit was amazing. They were inquisitive, interested and shared a desire to make a renowned service finer. Thank you one and all.

As to my supervisor, Rob Honey, I cannot thank him enough. He was a wonderful supervisor. Always there when I needed his support, and his advice and guidance was magnificent. He challenged my thinking and developed me from a firefighter to a researcher. He inspires me and I will be forever grateful.

For being allowed to spend so much time away from home I cannot thank my family enough. Their love and support have been constant and generous and materialised in so many wonderful, unexpected ways from loans of cars to surprise visits, and health advice to simple messages and phone calls. I love you all very much.

To my wife, Gita. Darling, your love has enabled me to do this. Without it I would have achieved nothing. It has helped me through my highs and lows by absorbing my grumpiness and frustrations and reflecting and sharing my success and happiness. It leaves me short of words to express how much that has all meant to me. Simply put, I love you, body and soul.

Publications

Chapter 2 is based on the published article: Butler, P.C., Honey, R.C., & Cohen-Hatton, S.R. (2020). Development of a behavioural marker system for incident command in the UK fire and rescue service: THINCS. *Cognition, Technology and Work, 22,* 1-12.

Chapter 4 is based on the published article: Butler, P.C., Bowers, A., Smith, A.P., Cohen-Hatton, S.R. & Honey, R.C. (2021). Decision making within and outside standard operating procedures: Paradoxical use of operational discretion in firefighters. *Human Factors* (in press).

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Context

In the 1990s a series of four research studies commissioned by the London Fire Brigade investigated the practices of incident commanders based on the principle that human factors were a critical feature of command for UK Fire and Rescue Service (FRS) incident commanders (Burke, 1997). Of these, the key study described by Burke (1997) produced a functional model of command that encapsulated the human factors involved (Figure 1). The model had two major components: deciding and acting and was developed via workshops involving subject matter experts in assessment or incident command competence. The deciding component received the essential information inputs necessary for incident commanders to make a decision, i.e. incident 'information', 'resources' and 'safety and risk'. These reflect the importance of situational awareness in decision making (see Klein, Calderwood, & MacGregor, 1989; Kaempf, Klein, Thordsen, & Wolf, 1996). The acting component represented the outputs of the commander. These were communication, control, and evaluation with the latter being recognised as generating more information that becomes another input into the deciding component. Overall, the model illustrates the importance of decision making for incident commanders and that communication and situational awareness are key functions of command.

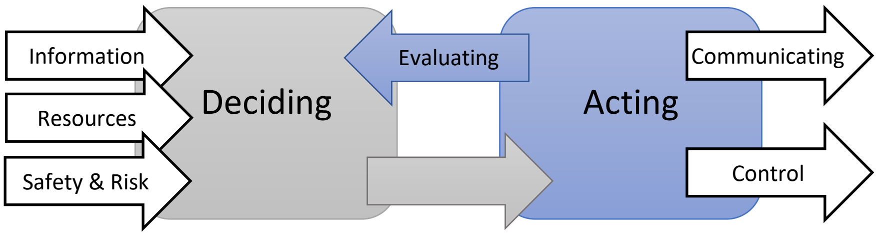


Figure 1: *Functional Model of Command (Adapted from Burke, 1997)*

The London Fire Brigade used the framework of the model to develop performance indicators for incident command, a database to gather evidence on performance to inform individual and organisational feedback, and units of competence for incident command (Burke, 1997). In turn this led to the development of an incident command model (Figure 2) and the introduction of training for non-technical skills, i.e., the 'cognitive, social and personal resource skills that complement technical skills and contribute to safe and efficient task performance' (Flin, O'Connor, & Crichton, 2008, p. 2). The training for incident commanders was akin to the aviation industry's crew resource management (CRM) training for flight crew that were first introduced in the early 1980s by airlines in the USA (see section 1.6 below; Helmreich, Merritt, & Wilhelm, 1999).

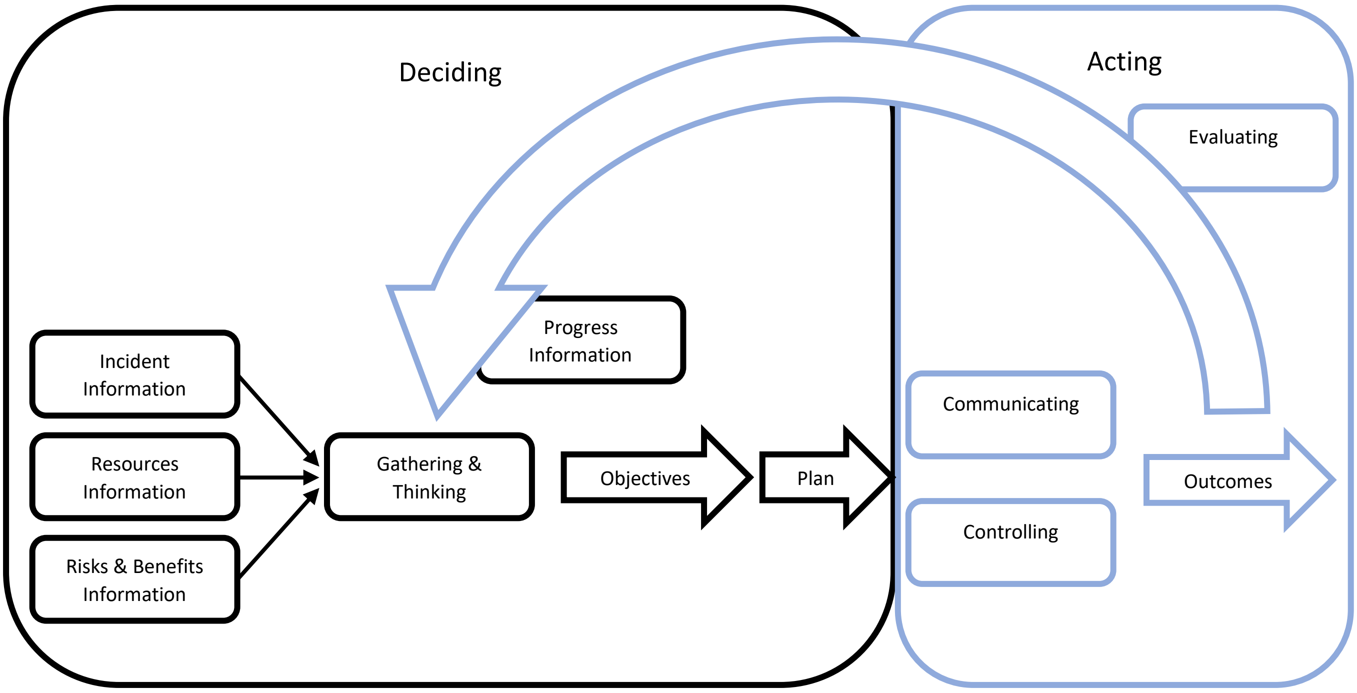


Figure 2: *Incident Command Model (Adapted from the London Fire Brigade, 2004)*

The research commissioned by the London Fire Brigade demonstrated an early recognition of the importance of non-technical skills for incident command. The decision making model and non-technical skills were later included in an appendix on the psychology of command by Professor Rhona Flin in the UK FRS national guidance for incident command (HM Government, 2008).

Between 2000 – 2010 the UK Fire and Rescue Service (FRS) has suffered the loss of 15 firefighters and FRS personnel while responding to emergencies (Fire Brigades Union, 2008; "List of British firefighters killed in the line of duty," 2021). In response to these losses the Health and Safety Executive (HSE; HSE, 2010a) in collaboration with the UK FRS published a set of principles. These set out to assist HSE inspectors and the public to understand how the UK's Fire and Rescue Authorities (FRA) could meet their requirements under health and safety law to protect firefighters and the public whilst delivering an effective fire and rescue service. Key amongst these, was the principle that FRAs had to ensure firefighters were prepared to make decisions in 'dangerous, fast-moving emotionally charged and pressurised situations, even when there may sometimes be incomplete or inaccurate information' (HSE, 2010a, p 2). With respect to HSE inspectors, they would consider the quality of firefighter decisions based upon the available information the firefighters had at the time of making them and whether the quality of those decisions reflected how adequately firefighters had been prepared. This established the importance of the decisions made under pressure for the firefighters in charge of emergencies[1] known as incident commanders.

[1] In the UK FRS the emergencies they respond to are referred to as incidents. Throughout the remainder of this thesis these terms may be regarded as interchangeable.

Over many years public inquiries and fatal accident investigations have identified a recurring need to improve incident commander decision making such as the 1988 Kings Cross Station fire (Fennell, 1988), 2007 London bombings (Hallet, 2011), 2008 Galston Mine incident (Torrie, 2012), and the 2009 Lakanal House high rise fire (Kirkham, 2013). Following a series of HSE inspections of eight FRSs in 2009/10 the HSE (2010b) made a number of recommendations about incident command and set out their expectations for arrangements FRSs should have in place with respect to incident commander decision making. Specifically, suitable training in relation to the judgement of risk and how to respond to and address unexpected events. The UK Government's Department of Communities and Local Government (DCLG; 2013) published an operational health and safety framework for FRAs with a comprehensive set of guiding principles that complemented the recommendations and expectations of the HSE (2010b). These included references to incident command, including leadership, risk management, competence, and operational procedures (e.g., the use of operational discretion to deal with unfamiliar incidents). Operational discretion is the use of professional judgement to make decisions outside of standard operating procedures (SOP) in order to save life, prevent serious escalation of the incident, or to prevent others taking action that may cause them harm (National Operational Guidance; N.O.G., 2021a). The framework featured the importance of human factors and recommended that FRAs should give consideration to the impact of such factors on incident outcomes, given their importance for effective safety. Further, the framework promoted the development of non-technical skills as means to positively influence safer firefighter behaviour at an incident and highlighted the adverse impact stress and fatigue can have on individual performance, including decision making (DCLG, 2013). Core non-technical skills include leadership, decision making, managing stress and coping with fatigue (Flin et al., 2008). The importance of the use of operational discretion and professional judgement was reiterated when the Chief Fire Officers Association (now the National Fire Chiefs Council; 2015) identified it as an area in need of critical attention along with incident commander decision making. The report also stated there was a lack of research into incident command, especially human factors-related research and highlighted how other safety critical industries had developed behavioural markers systems to evaluate the performance of non-technical skills, including decision making.

Recent research has involved examinations of non-technical skills. Butler (2016) identified the non-technical skills of UK FRS incident commanders as: assertive, effective and safe leadership, decision making and planning, inter-personal communication, personal resilience, situational awareness[2], and teamwork and interoperability. A feature of that research highlighted that all 27 FRSs who responded to a survey (approximately 56% of the 48 surveyed out of 53 UK FRSs) indicated that incident commanders received decision making training. 70% or more of the FRSs provided incident commanders with training in communication, teamwork, situational awareness and leadership. However, this reduced to just over half for the provision of theoretical training related to personal resilience, which further reduced to a third for practical training.

With respect to decision making, Cohen-Hatton, Butler, and Honey (2015) reported that tactical, on scene incident commanders made intuitive decisions irrespective of whether it was appropriate to do so in a given situation, i.e., routine or complex. This was at odds with the extant decision making model (Figure 2). This finding was replicated by

[2] Situation awareness and situational awareness refer to the same concept and as such these terms may be regarded as interchangeable throughout this thesis.

Wilkinson, Cohen-Hatton, and Honey (2019) for remote, strategic commanders forming multi-agency, strategic coordination groups, with respect to their application of the Joint Decision Model (Joint Emergency Service Interoperability Principles; JESIP,2016).

Decision making closely linked to situational awareness. Situational awareness and situation assessment are at the forefront of many decision making models, for example, the Recognition-Primed Decision model (RPD; Klein, 1993) and the general decision making model (Flin et al., 2008). Flin et al. (2008) demonstrated the importance of the relationship by describing situation awareness as '…the cognitive processes for building and maintaining awareness of a workplace situation or event.' (Flin et al., 2008, p. 17). They highlight how situation assessment is a critical feature of situation awareness because it is triggered by changes detected in the environment being monitored. Newell, Lagnado, and Shanks (2015) also demonstrate the importance of situation awareness when making decisions in dynamic environments. They describe the critical roles of information gathering, understanding information, and then using information (feedback) to judge the effectiveness of a decision based on anticipations to learn from its outcomes. In a follow-up to their earlier decision making study, Cohen-Hatton and Honey (2015) investigated the affect of goal-oriented training on the decision making processes of incident commanders, i.e., situation assessment, plan formulation, and plan execution. The goal-oriented training required incident commanders to use a set of decision controls whenever they made a decision. The decision controls were a rapid mental check of their goals, the anticipated consequences, and an evaluation of risks versus benefits. They were a mechanism to ensure decisions were routinely evaluated. Cohen-Hatton and Honey (2015) found a shift in the decision making processes used by incident commanders from situation assessment to plan execution to situation assessment to plan formulation. This was accompanied by significantly more anticipatory situational awareness in incident commanders exposed to goal-oriented training compared to those who received standard decision making training.

It was decided by the author (Philip Carl Butler, PCB) to initially develop and evaluate a behavioural marker system based upon the non-technical skills identified by Butler (2016) which would include an examination of the relationships between them. This would be followed by a study influenced by the findings from the recent UK FRS studies referred to above because they warranted further research to learn more about the relationship between decision making, situational awareness and personal resilience.

Having established that situational awareness and decision making are non-technical skills and that decision making under pressure (e.g., in the context of using operational discretion), is an area of concern to the UK FRS, the next section in this chapter will outline a brief history of decision making research and theory followed by sections that briefly review situational awareness and the nature and influences of (acute and chronic) stress. The chapter will proceed by describing the incident command context in which incident commanders make decisions using an example to illustrate decision making under pressure. This will be followed by a brief, general history of non-technical skills and behavioural marker systems. Finally, the objectives of this thesis will be presented alongside the methodological approach used to address them.

1.2 A brief history of individual decision making

Life is the art of drawing sufficient conclusions from insufficient premises.

Samuel Butler (1912)

Decision making may be defined as a commitment to a course of action to achieve goals (Yates, Veinott, & Patalano, 2003). It is a fundamental aspect of human behaviour and is the process of selecting an option with respect to the prevailing situation (Flin et al., 2008). The decision making process brings together cognition and emotion as people apply rationale and emotional values to assist them through it (Bechara, Damasio, & Damasio, 2000). Fellows (2004) described the basic decision making process as one involving option generation, evaluation and choice; and Hastie and Dawes (2010) highlight how the process of decision making is founded on an individual's situational awareness. Two or more options are generated to which individuals attach expectations about what might happen and what the outcomes will be based on an individual's beliefs or probabilities. This indicates the inherent ambiguity about how uncertain future events will influence the nature and outcomes of a decision (Hastie, 2001). The positive or negative consequences of the outcomes are evaluated based on the individual's values and goals prior to a choice being made (Hastie & Dawes, 2010). Most psychological decision making models in the literature have this as a common pathway through the decision making process, but they are only part of the history of decision making theory and research.

1.2.1 The origins of modern decision making theory

In 1494 a Franciscan monk called Paccioli published a book on mathematics which, amongst other things, considered how to resolve a gambling problem of what proportion of the stake to give each player when a game is interrupted (Almy & Krueger, 2013). Pacciolo did not resolve this problem, instead it was Pascal and Fermat who did so some 160 years later (Almy & Krueger, 2013). They determined that to divide the money up fairly the predicted or expected value of each gamble had to be calculated. For example, if the game was to win five gambles to collect £100, how do you divide the money up when one player has won on four occasions and the other just once when the game has to stop? Intuitively, you might think the divide should be 80% to 20%, but when you consider the expected values of each gamble you can see that this is incorrect. Expected values are calculated by multiplying the value of the money by the probability of a win occurring. In the example given each gambler has a 50% chance of winning, for the trailing gambler they would need to win the next four gambles to win the game: £100 (0.5 x 0.5 x 0.5 x 0.5) = £6.25. Therefore, the leading gambler should receive 93.75% of the stake £93.75. This probability-based decision represents a rational decision because each player gets a share of the stake based on their individual expected value at the time the game stopped (i.e., what it was worth to them; see Almy & Kreuger, 2013). As such, this solution represents an underlying principle that decisions are made by individuals to maximise expected value. A further development in decision making research can be attributed to Bernoulli (1738) who took account of the effect of emotion in decision making and proposed replacing the concept of expected value with that of utility (i.e. pleasure or usefulness). He recognised that the utility of money declines the more you have. For example, if you are a billionaire then winning £1,000,000 does not increase your wealth by much, but if you earn the average Great British salary of £29,848 (Office of National Statistics, 2021) then winning that amount has a large utility value.

Approaches to decision making can be divided into two broad categories: descriptive and prescriptive. Descriptive models explain the process of decision making whereas prescriptive models identify the requirements for making good or correct decisions under normative standards (Baron, 2000). Each type of model has developed our understanding of decision making and the contribution of key models from each approach will be briefly described in the next sections beginning with the prescriptive model of Expected Utility Theory (Von Neumann & Morgenstern, 1947).

1.2.1.1 Expected Utility Theory. Expected Utility Theory (EUT) originated from the field of economics. Von Neumann & Morgenstern (1947) stated that decision makers seek to maximise utility; that is, a subjective value placed on an outcome. When confronted with the need to make a choice they assess the expected utility (value) of each option by multiplying the probability of a given outcome with the utility of the outcome. They introduced axioms with which to assess option preference based on the principle of maximising expected utility. The four that define a rational decision maker are:

- Completeness (preferences are well defined so an individual can always decide between any two alternatives)
- Transitivity (an individual decides according to the completeness axiom and choses consistently, e.g., if an individual prefers option 1 over option 2 and option 2 over option 3, then they will prefer the option 1 over option 3)
- Continuity (an optimising gamble option is preferred to a certain gain)
- Independence (an individual decides according to the completeness axiom preference so, if option 1 is better than option 2 that will remain the case upon the introduction of an irrelevant option 3) (Von Neumann & Morgenstern, 1947)

Savage (1954) added the element of subjectivity to the theory by proving that whilst satisfying all the axioms, individuals assign subjective probabilities or weights to the possible outcomes of an option. Yet, do the rules of EUT really describe what decision makers do? Do people always follow the rules of EUT? In short, they do not appear to. Many studies have found that the rules do not appear to be adhered to when decision makers are operating in uncertain environments (see Goldstein & Hogarth, 1997). Decisions made in accordance with EUT are made under conditions where the probabilities for each option can be calculated and rules applied in order to select the best one. However, firefighters work in the realms of uncertainty, unpredictability, and sometimes the unknown, where information is missing and the situation is dynamic and unstable. Such complex decision making environments are not conducive to rational decision making, so it has been argued that prescriptive approaches have limited use in understanding real-world decision making (Simon, 1956). Rather the decisions that are made are influenced by the environment.

1.2.1.2 Bounded Rationality. Bounded Rationality is a descriptive model of decision making that describes how decision makers are constrained by the information they have, their cognitive limitations, and the time available to make a decision (Simon, 1956). In essence, the decisions they make are as rational as they can be based upon the context in which they are made and the need to produce workable solutions to problems. This is achieved by decision makers using strategies that shortcut the rational decision making process (e.g., estimates). Simon (1956) emphasises that what matters is how well matched the mind of the decision maker is with the environment rather than maximising expected utility. He described bounded rationality as a satisficing heuristic where a decision maker considers options one at a time until finding an adequate one (Simon, 1978). Tversky's (1972) elimination-by-aspects theory is similar to Simon's approach. Through a process of eliminating options by assessing one aspect at a time until only one option is left. For example, consider the purchase of a car. A number of aspects would be considered including the make of car, its impact on the environment, cost, rate of acceleration from 0-60 mph, and so on. So each time a car does not possess one of the desired aspects it is eliminated, progressively reducing the number of options. However, this approach is limited because a different option remains at the end based on the order in which these aspects are considered and consequentially the best choice may not be made. However, as we shall see next Tversky and Kahneman (1974) went on to explore and develop the concept of bounded rationality and the use of heuristics to make decisions in uncertain environments.

1.2.1.3 Heuristics and biases. Tversky and Kahneman (1974) investigated human judgement to explain why people make errors under conditions of uncertainty. They argue that reliance on the use of heuristics is prevalent in decision makers because they can be used very quickly and limited demands are placed upon cognitive processing. However, errors occur because the heuristic approach ignores, for example, base rate information (i.e. 'the relative frequency with which an event occurs or an attribute is present in the population'; Koehler, 1996, p. 1). Tversky and Kahneman (1974) highlighted three heuristics that influence or bias decisions. The representativeness heuristic is used when judging the likelihood of an event belonging to specific category or process (e.g., incorrectly attributing an occupation to an individual based on a description of them). Judgements can be biased because comparisons are rapidly made with existing mental stereotypes. The availability heuristic is used when judging the frequency of an event and is biased by how easily something is retrieved from memory: The easier it is to recall something, the more influence it has. People use the anchoring heuristic when making a numerical estimate in light of an existing number, which biases their judgement by influencing how far their estimate moves away from it (e.g., when asked to value a property after someone else has provided a valuation). Todd and Gigerenzer (2007) also suppose that in the face of complex and challenging situations decision makers make use of heuristics. They argue that fast and frugal heuristics are based on a one-reason principle, such as the 'recognition' heuristic, where people select the only one of two options that they recognise. As the name suggests people use fast and frugal heuristics when needing to make rapid decisions because they are easy and take up little cognitive processing power.

In keeping with the concept of biases and making decisions under conditions of uncertainty, Kahneman and Tversky (1979) developed Prospect Theory. The theory deals with gain and losses and how people respond to both when making a decision. It addressed some of the short-comings of EUT and centred on three cognitive concepts that are vital in the evaluation of outcomes: a reference point, a principle of diminishing sensitivity, and loss aversion (Kahneman, 2011). When making a decision individuals have in mind a reference point, such as the status quo or something they might expect from a given situation. For example, when selling your home, you have in mind the price you expect to sell it for. Any outcome that is better than the reference point is seen as a gain, whereas anything less is seen as a loss. The principle of diminishing sensitivity refers to the subjective weighting applied by individuals to losses and gains, which is greater for losses than gains so people are risk averse toward gains and risk seeking towards losses (Kahneman, 2011). Finally, loss aversion, where Kahneman and Tversky (1979) found that people have a greater sensitivity to losses than to gains. For example, the fear of losing £50 is felt much more than the prospect of winning £100 on the toss of a coin. The involvement of emotion in decision making is also evident in neurological studies. Shiv, Loewenstein, Bechara, Damasio and Damasio (2005) conducted a study comparing the performance of patients with brain damage in areas associated with emotion (e.g., amygdala, orbitofrontal cortex) with that of a group of patients with damage to other parts of the brain, and a group of healthy controls. They asked the participants to invest $1 on the toss of coin where they would gain $1.50 if it was tails or lose their money if it was heads. On average, each coin toss could net the participant with a 25 cents gain, so to maximise their profit taking part in every coin toss was the best strategy. They found that patients with brain damage to the areas of emotion invested on 84% of the coin tosses compared to 61% for those with damage to other parts of the brain and 58% for the healthy controls (Shiv et al., 2005).

To briefly summarise, decision making research and models have developed from gambling studies in the 1400s that led to rational, probability-based models such as EUT to those that take account of cognitive limitations and the environment as emphasised by Bounded Rationality. Under conditions of uncertainty people use heuristics and are influenced by biases to make decisions because they are quick and take up little cognitive capacity to make, but they are subject to error as often base rate information is ignored. Together, Bounded Rationality and the use of heuristics take account of the impact of a hostile environment and human limitations might have on the decision making of an incident commander and as we shall see, they form the foundations of Naturalistic Decision Making.

1.2.1.4 Naturalistic Decision Making (NDM). NDM is founded on the idea that experience and expertise affect decision making (Klein, Shneiderman, Hoffman, & Ford, 2017). NDM research emerged in the 1980s because laboratory-based research had demonstrated that people did not make rational, optimal decisions as evidenced by their use of heuristics and biases (Klein, 2008). NDM is defined as 'the way people use their experience to make their decisions in field settings' (Zsambok, 1997, p. 4). The term 'field settings' refers to complex, ambiguous, and challenging environments (Klein, 1993). Klein (1993) conducted a study of FRS incident commanders using cognitive task analysis, a method that uses semi-structured interviews to explore a specific event recalled by the participants. He found a characteristic of the FRS incident commanders' decision making was the absence of a rational decision making process to maximise utility. Rather, incident commanders made very rapid decisions in direct response to the (changing) environment. This led to the development of the RPD model (Klein, 1993; 1998) which has three levels that are influenced by the prior experience of decision makers (Lipshitz, Klein, Orasanu, & Salas, 2001). The first level involves an assessment of the situation and a recognition of what action to take. The second level relies on building a mental model of how the incident developed to its current status when the situation is not immediately obvious. Finally, the third level uses mental simulation to imagine how a course of action will impact the incident. In order to make decisions in this way requires an individual to have experience to draw upon in order to recognise situations, create mental models and generate mental simulations (Lipshitz et al., 2001). Individuals with the ability to do all of these are regarded as experts, and research with high reliability organisations (e.g., the military and commercial pilots) confirmed RPD as the most used decision making strategy (Lipshitz et al., 2001). More recent research using video, rather than recollection, has found evidence consistent with the use of RPD in incident commanders in the UK FRS (Cohen-Hatton et al., 2015; Cohen-Hatton & Honey, 2015).

In fact, the use of RPD is in some ways embedded in the UK FRS through the provision of standard operating procedures (SOP) for familiar types of emergency incident. SOPs inform safety critical personnel how to organise themselves and conduct specific operations (Brunacini, 2002). As such, they provide sufficient guidance for them to deliver a standard, consistent response (LaPorte & Consolini, 1991). However, SOPs have obvious limitations. Klein (2009) lists these as being insufficient, especially in a complex and novel situations, and being difficult to maintain inducing people to be complacent and blindly follow them, which limits the development of expertise. Indeed, this issue has been recognised in the UK

FRS, where National Operational Guidance licenses the departures from SOPs, and the use of operational discretion, when there is a danger that an incident will escalate rapidly or there is risk to human life (N.O.G., 2021a). Of course, these conditions are likely to induce acute stress, which will itself interact with the processes involved in decision making. One objective of this thesis is to examine how different types of incident affect stress and influence the process of decision making (see Chapters 4 and 5).

1.3 A brief overview of situational awareness

The concept of situational awareness emerged during the First World War in relation to military aviation when it was regarded as important for military aircraft crews (Press, 1986; cited in Endsley, 1995). Yet why is it important? A recent meta-analysis of 46 studies across a wide range of different domains found that situation awareness was predictive of good performance in 89% of the studies (Endsley, 2021). However, it is clearly not the only factor involved, and while good situation awareness enhances performance it does not necessarily guarantee good outcomes (Endsley, 2021).

A widely accepted definition of situation awareness is ‘the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and projection of their status in the near future’ (Endsley, 1995, p. 36). More broadly, situational awareness refers to the cognitive processes that enable an individual to build and maintain awareness about an environment or event (Flin et al., 2008), or representing a mental model of the current situation that forms the central organising feature for all decision making (Endsley, 1997). Perhaps the most influential model of situational awareness is the three-level model proposed by Endsley (1995; for a review, see Salmon, Stanton, Walker & Jenkins, 2009). It is cited in the UK FRS (N.O.G., 2021b) and will be the focus of interest here.

The three-level model developed by Endsley (1995) describes the cognitive processes that underlie situational awareness and emphasises the dynamic nature of it. The three hierarchical levels are perception, comprehension, and projection. Level 1 situation awareness (perception) refers to perceiving information relating to, and therefore being aware of, the attributes and dynamics of various objects in an environment. Level 2 (comprehension) refers to the integration and interpretation of that information to understand what is happening in that situation. Levels 1 and 2 are influenced by individual goals, objectives, preconceptions about the situation, expectations and experience in the form of mental models. Level 3 (projection) involves the prediction of a future state based on the knowledge acquired via the first two levels and the mental models of previous experience (Endsley, 1995).

The three levels within the model form part of the information processing conducted to make a decision and take action. They are also influenced by individual, task and systemic factors such as training and workload, level of complexity, and interface design, respectively. One of the key assumptions of this model is the critical role mental models play in the building and maintaining of situation awareness. Endsley (1995) advocates that environmental cues are matched to mental models, which in turn facilitate the development of situation awareness. The mental models formed by training and experience direct attention to the cues (Level 1), guide the integration of them to aid comprehension (Level 2), and generate potential future states (Level 3). This suggests that experienced incident commanders will build and maintain situation awareness more effectively than inexperienced incident commanders.

1.4 A brief overview of stress

Stress may be described as chronic (persisting over a long period of time) or acute (lasting a few seconds or hours; Flin et al., 2008). It affects the physiology, psychology and behaviour of people, impacting performance and health (for a review, see Staal, 2004). Research into the impact of physiological stress began in the 1930s (Selye, 1956), but it has since become the province of psychology, with a burgeoning literature (Mason, 1968) including work that has had significant impact in occupational psychology (e.g., Lazarus & Folkman, 1984). Cox (1978) described three models of stress: (i) the engineering model (external stimuli cause a stress reaction), (ii) the physiological model (the body's internal response to stress), and (iii) the transactional model (individual perception of demand versus ability to cope). Although there are different models of stress it is generally accepted that responses to stress include the physiological, psychological and behavioural. So what effects does stress have? Selye (1956) developed the concept of General Adaptation Syndrome (GAS) based on his observations of how ill people seemed to share the same characteristics irrespective of the illness (e.g., loss of appetite and ambition). GAS encapsulates the body's response to stress with respect to its nervous system, the sympathetic adrenal medullary (SAM) response system (that produces adrenaline to initiate the flight or fight response), and endocrine system, the hypothalamic-pituitary-adrenal (HPA) axis that produces corticosteroids, (including cortisol to sustain the stress response; Webster-Marketon, & Glaser, 2008). Another corporeal response triggered by the endocrine system is that of the immune system which uses leukocytes to tackle stress or disease. A brief description of the acute response of leukocytes is provided in Chapter 4, where a measure of leukocyte function is used in the context of incident commanders responding to simulated incidents. Once an acute stressor has been removed, the body returns to its normal homeostatic equilibrium. However, if such stressors are not removed then the resulting imbalances can cause illness such as heart disease or inflammatory diseases (Mariotti, 2015).

Another feature of stress is that there are individual differences people show in response to the same stressor (see Kudielka, Hellhammer, & Wüst, 2009 for a review). For example, Kiecolt-Glaser et al (1995) chronic stress study on carers found that wounds caused by a harmless biopsy procedure took significantly longer to heal for carers compared to a control group. Likewise, with respect to acute stress, people's ability to cope will differ depending on how they appraise the situation and their ability to cope with it (e.g., Lazarus & Folkman, 1984). It is difficult to define the concept of stress in a way that unifies these distinct influences and takes account of individual differences (Stokes & Kite, 2001). However, a review of stress research by Koolhaas et al. (2011) defined it as occurring when demands upon an organism exceed its regulatory capacity, especially when circumstances are unpredictable or uncontrollable (see also, Lazarus & Folkman, 1984).

The relevant research on how stress impacts decision making processes in laboratory settings is considered in more detail in Study 3a. Of more immediate relevance here is that accident investigations have suggested that acute stress has an impact on the performance of non-technical skills to jeopardise safety (Crichton, Lauche, Flin, 2005; Flin et al., 2008; Weick, 1990). The HSE guide (1999) 'Reducing error and influencing behaviour' illustrates a range of circumstances wherein human error is more likely. The categorised stressors were environmental (heat, noise, etc), task-related (overload, distractions, interruptions, etc), organisational (inter-personnel conflict, peer pressure, etc), individual (fatigue, inexperience, family problems, etc), and equipment (e.g., poorly designed displays, confusing SOPs). The DCLG (2013) health, safety and welfare framework for the operational environment states that managing task demands is a way of managing the likelihood of an error occurring leading to an accident. It also describes how fatigue affects an individual's cognitive functions, including their judgement, perception and attention (conclusions that are supported by laboratory research; see Chapter 4). In a review of organisational resilience, Crichton, Ramsey and Kelly (2009) identified several recurring themes from lessons learned following incidents, drawn together across a range of different industries. These included needs to manage the risk of fatigue and to develop the non-technical skills of those responding to the event under conditions of uncertainty and stress. As we have already noted the core set of non-technical skills includes managing stress and coping with fatigue (Flin et al., 2008; see also, Brunsden, Hill & Maguire, 2014).

1.5 UK Fire and Rescue Service incident command

The UK FRS respond to a wide variety of emergencies or incidents (as they are known in the service) that comprise fires, rescues, and hazardous materials that involve industry, various types of building, transport, and utilities and fuel infrastructure (N.O.G., 2021c). Incident commanders are responsible and accountable for the command and control of the emergency response to an incident, the identification of the hazards and risks involved, and the determination of a plan and resources required to implement it (N.O.G., 2021d). To assist with this, incident commanders use an incident command system (ICS) to manage resources which allows them to create a hierarchical network of supporting officers and teams to take responsibility for operational tasks and functions (N.O.G., 2021e). In support of the safe and efficient resolution of incidents, the ICS is designed to ensure incident commanders create visible lines of command and maintain manageable spans of control whilst delegating authority and creating a communications network (N.O.G., 2021e). Additionally, FRS incident commanders work closely with those from other agencies. The Joint Emergency Services Interoperability Principles (JESIP; 2016) stipulate they are expected to co-locate and communicate effectively to coordinate a multi-agency response and ensure the other agencies have a shared understanding of the risks and situation.

National Operational Guidance (N.O.G., 2021f) details the four levels of UK FRS command: initial, intermediate, advanced, and strategic. Initial level incident commanders are supervisory level officers responsible for resolving routine incidents or the early stages of larger incidents. Intermediate level incident commanders are middle managers that take charge of medium-large incidents requiring greater command and control. Advanced level incident commanders are senior officers who take charge of very large and complex incidents. Strategic commanders are principal officers who operate remotely from the scene of an incident and do not take over tactical command of the scene. Strategic commanders take command of a multi-agency strategic coordinating group focused on long-term incident response and recovery (N.O.G. 2021f). For each level of command there is a corresponding qualification. The qualification units are comprised of a series of learning outcomes to be achieved that broadly relate to the incident command system, the roles of others, decision making, leadership, risk management, planning and debriefing (for details see Skills for Justice, n.d.). For each managerial role within the UK FRS, i.e., from supervisory managers to brigade managers, there is a national occupational standard with an incident command unit of competence appropriate to the level of incident command required to be performed by the role (Skills for Justice, 2013a; 2013b; 2014). These are:

- Lead and support people to resolve operational incidents (for initial level incident commanders)
- Lead, monitor and support people to resolve operational incidents (for intermediate and advanced level incident commanders)
- Provide strategic advice and support to resolve operational incidents (for strategic level incident commanders; for details see Skills for Justice, 2013a; 2013b; 2014)

The national occupational standard for intermediate and advanced level incident commanders requires them to competent at: reviewing and determining the status of an incident, assuming responsibility and implementing action to support those involved in the incident, and debriefing personnel after an incident (Skills for Justice, 2013b). The variety of factors that underpin an incident commander's competence are wide ranging. For example, reviewing and determining the status of an incident requires that an incident commander is able to:

- Gather information on progress, risks, resources (including other agencies) and incident management, including:
    - The priority actions
    - The risks and implications for personnel, the community and for the wider context, e.g., businesses, public transport and healthcare services
    - The roles and responsibilities of key personnel and how to contact them

- Identifying and resolving any discrepancies between pre-incident and at-incident information
- Confirming that current action complies with relevant legislation and protocols, e.g., health and safety and standard operational procedures
- Anticipating future resource needs (Skills for Justice, 2013b).

All of this is undertaken in dynamic, changing, and dangerous working environments that may be unfamiliar and enveloped by unrealistic public expectations of what firefighters can achieve (HSE, 2010a). These characteristics align with those of high reliability organisations (Roberts & Rousseau, 1989). For example, firefighters operate in very complex working environments with multiple decision makers in complex communications networks, and time poor situations in which rapid and critical decisions must be made. Therefore, it is reasonable to assume that UK FRSs are appropriately regarded as high reliability organisations. Such organisations are defined as being able to deliver almost error-free performance despite the hazardous working environments where errors could have catastrophic consequences (Lekka, 2011). Such errors become the subject of accident investigations and where necessary public inquiries[3] that identify the failures that culminate in errors being made.

1.5.1 A recent example: The Grenfell Tower fire

On the 14th June 2017 at 00:59 hours the first London Fire brigade incident commander arrived at Grenfell Tower in response to a report of fire in a flat on the 4th floor of the residential block of flats (Grenfell Tower Inquiry, 2019). The Grenfell Tower Inquiry

[3] Between 2005 and 2018 the UK Government spent £239 million on 26 inquiries initiated and concluded during that time (National Audit Office, 2018)

(2019) reports that 10 minutes later and just before the first firefighting crew enters the flat, the fire breaks out of a window (Figure 3) and sets fire to the exterior cladding on the building. It takes just 18 minutes for the fire to travel from the 4th floor to the roof of the 67m, 24 story building (Figure 4). Four minutes after the fire reached the roof the incident commander requested 25 fire engines and they remained in charge of the incident for another 19 minutes, totalling almost 1 hour. The first incident commander was a an initial incident commander, i.e. a supervisory officer able to take charge of small and developing incidents (N.O.G., 2021f). By the time they handed over command there were 11 fire engines and five specialist vehicles in attendance totalling 60 or more personnel (Grenfell Tower Inquiry, 2019). This level of resources is redolent of a large and serious incident that would typically be managed by a Level 3, advanced incident commander (N.O.G., 2021f).

Figure 3: *Image of Fire Breaking Out of 4th Floor Window 10 Minutes After Arrival of First Incident Commander*

Figure 4: *Image of Fire Reaching the Roof of the 24 Story Building 18 Minutes After Breaking out of 4th Floor Window*

The Grenfell Tower fire presented the first incident commander with challenges that they admitted were beyond their ability to cope. For example, they reported they had no experience to rely upon, their senses were overload, they had feelings of helplessness, and of being unable to cope with the relentless and rapid development of the incident (Grenfell Tower Inquiry, 2019). It is also reasonable to infer that the incident commander was unfamiliar with the situation and misperceived the risk. Each one of these factors is an error-producing situation, where dealing with a novel situation without any insight into the potential consequences carries the highest probability of error (Reason, 2016). Reason (2016) describes three grounds for people making errors when engaged in novel problem-solving. Two relate to human information processing (Wickens, Lui, Becker, & Lee, 2004): The span of attention and working memory, and the creation of an inaccurate mental model within your mind's eye; and the third involves emotion and, in particular, stress. The limited capacity of working memory leads to forgetting, so once its capacity is reached as attention shifts to another aspect of the problem an earlier memory is forgotten, which in turn results in the development of an incomplete mental model. With respect to stress, Reason (2016) focuses on how it influences people to resort to familiar courses of action. It is reasonable to assume that the first incident commander of the Grenfell Tower fire was stressed, which is supported by one of the key findings of phase one of the Grenfell Tower Inquiry (2019) that stated none of the early incident commanders were able to make the decision to cease using the standard 'stay put' policy for residents and switch over to a mass evacuation of the building. These incident commanders followed the familiar course of action when confronted with a situation that required them to make decisions beyond the scope of the SOP.

This description of the experiences of incident commanders involved in the Grenfell Tower fire highlights the role of what are referred to as non-technical (or psychological) skills that impacted the response to managing and resolving the incident. The UK FRS has national operational guidance on non-technical command skills and advocates that incident commanders are trained in their use, unlike other high reliability industries it does not have a means to evaluate their use (e.g., a behavioural marker system).

1.6 A brief history of non-technical skills and behavioural marker systems

In the early 1980s, commercial USA airlines were the first high reliability organisations to recognise the importance of non-technical skills to flight safety and to develop non-technical skills training programmes (Helmreich et al., 1999). This was an outcome of a 1979 National Aeronautic Space Agency (NASA) conference, which reported research into a series commercial aircraft accidents that attributed their cause to pilot error, and specifically failures of interpersonal communications, decision making, and leadership (Helmreich et al., 1999). To address these interpersonal issues research was undertaken to identify the non-technical skills that contributed to safe flight by analysing accident reports, interviewing pilots and running experiments on flight simulators (Kanki, Anca, & Chidester, 2019). Once the non-technical skills were identified, training was developed to improve error management to make for safer flight operations. Non-technical skills training became known as CRM (Helmreich et al., 1999). Initially CRM training was standalone training focused on the individual, but over time it has evolved to become more integrated into existing aircrew training with a team focus (Kanki, et al., 2019). CRM is defined as using all available resources to fly safely and efficiently (Lauber, 1984) and embodies the human factors of aviation. The simple relationship between non-technical skills and safety is that good non-technical skills practice makes errors less likely to occur (or if they do they are more likely to be caught and addressed) and fewer errors reduces the likelihood of an accident occurring (Flin et al., 2008), resulting in safer operations.

Since 2000 other high reliability industries have adapted CRM training programmes to suit their own needs (see Flin, O'Connor & Mearns, 2002; Salas, Wilson, Burke & Wightman, 2006 for reviews; see Schuermann & Marquardt, 2016 for lessons learned), including the fire and rescue service (Okray & Lubnau II, 2004). However, as the frequency of accidents within the aviation industry is extremely low, judging the impact of CRM training on improvements in flight safety would unfeasible (Flin et al., 2002). One of the alternatives was to observe aircrew behaviours and measure their non-technical skills performance using behavioural marker systems (Flin et al., 2002).

The development of behavioural marker systems was championed in the USA with the first system produced for pilots in the late 1980s based on the results of a research project aimed at validating the transfer of CRM training from the classroom to the cockpit (Flin et al., 2008). Klampfer et al. (2001) define a behavioural maker system as a 'taxonomy or listing of key nontechnical skills associated with effective, safe job performance in a given operational job position…with some decomposition of major skill areas usually illustrated by exemplar behaviours' (Klampfer et al., 2001, p. 7). Besides the non-technical skills (and their sub skills) and behavioural makers, a behavioural marker system also includes a rating scale (Flin et al., 2008). Progressively over the next 30 years many non-technical skills training and behavioural marker systems were developed for safety critical personnel within high reliability industries. What follows is a selective history of the introduction of behavioural marker system into some of those industries.

In 1996 the European aviation regulator initiated a project to evaluate the non-technical skills of aircrew that culminated in the development of the NOTECHS behavioural marker system (Flin et al., 2003). More recently systems have been developed for offshore transportation helicopter pilots and for search and rescue helicopter pilots (Hamlet, Irwin, Flin, & Sedlar, 2020a; 2020b). The first behavioural marker system within healthcare was developed in 1998 (Gaba et al., 1998) for anaesthesiologists. Since then several systems have progressively been developed including anaesthetists (ANTS; Fletcher et al., 2004), surgeons (NOTSS; Flin et al., 2006), scrub nurses (SPLINTS; Mitchell et al., 2013), and anaesthetic nurses (ANTS-AP; Rutherford, Flin, Irwin, & McFadyen, 2015). The maritime world saw the development of a behavioural marker system for the merchant navy (Devitt & Holford, 2010), a prototype system for US Navy officers of the deck (O'Connor & Long, 2011) and one for Portuguese rating cadets (da Conceição, Basso, Lopes, & Dahlman, 2017). The UK Railway Safety and Standards Board (2012) have introduced a behavioural markers system for train drivers and within the oil and gas industry, a behavioural maker system for drilling teams has been developed (Crichton, Moffat & Crichton, 2017) with a recent industry report advocating the expansion of non-technical skills training and the use of behavioural marker system across other areas of oil and gas operations (International Association of Oil and Gas Producers, 2018).

Behavioural rating systems have been shown to improve safety (McCulloch et al., 2009; see Kodate et al., 2012 for a review) and are valued as an assessment and research tool (Flin et al., 2008). However, while the non-technical skills for incident commanders were identified for firefighters in the USA (Okray & Lubnau II, 2004), a behavioural marker system was not subsequently developed. In the UK, national operational guidance (N.O.G., 2021g) describes the non-technical command skills that should be used by incident commanders, but a system to assess the quality of their use has not been developed. The principal objective of this thesis was to develop and validate such a system, and to use it as one basis for evaluating the impact of stress on the response of incident commanders to different types of incident.

1.7 Objectives and methodological approach

The objectives of this thesis were to:

1. Develop a behavioural marker system for incident command in the UK Fire and Rescue Service.
2. Evaluate the resulting behavioural marker system.
3. Use the system to understand how non-technical skills, in particular decision making, situational awareness and personal resilience are deployed in different types of incident.

Briefly, to meet the first objective, a series of workshops were facilitated with subject-matter experts. These workshops allowed the behavioural markers for an identified set of non-technical skills (Butler, 2016; see also, Butler et al., 2020) to be determined alongside the rating scale for the system (Chapter 2). The second objective was achieved through a preliminary evaluation of the system (Chapter 2) and a more extensive evaluation (Chapter 3). This evaluation involved assessing the reliability, sensitivity, usability, and validity of the system, through a long-term evaluation involving seven UK FRSs; and secondary analysis of archival data. The final objective was achieved in a large-scale study examining the use of SOPs and operational discretion by incident commanders, which also included the presence of acute stress.

Chapter 2

Study 1: The Development of a Behavioural Maker System for UK Fire and Rescue Service Incident Commanders

2.1 Abstract

The performance of command skills[4] by incident commanders' impact upon firefighter health, safety and welfare. However, in the UK FRS only a prototype set of such skills existed that underpinned incident command and the service did not have a system in place to assess those skills. The aim of Study 1 was to produce a system able to measure the performance of command skills in the assessment, training and operational environments of the UK FRS; and to facilitate future research. The objectives of this research were to develop a behavioural marker system and to refine the prototype set of command skills. Subject matter experts drawn from the UK FRS participated in a series of workshops to achieve the project's aim and objectives. The prototype set of command skills comprised of: Assertive, effective and safe leadership; effective decision making and planning; interpersonal communication; personal resilience; situational awareness; and teamwork and interoperability. Exemplary behaviours were identified for each of the sub skills underpinning the command skills and necessary amendments to the skill set made. A rating scale and means to capture and feedback on performance were also developed to complete the behavioural marker system. The general profile of the prototype command skills was similar to those for other high reliability industries, their precise nature differed by dint of the context in which they occur. The resulting behavioural marker system, The Incident

[4] 'Command skills' is the term used in the UK FRS for non-technical skills. Throughout the remainder of the thesis the two terms may be regarded as interchangeable.

Command Skills (THINCS) system, provides the UK FRS with the means to monitor and assess the performance of command skills by incident commanders in assessment, training, and operational environments. It also provides a means to validate the transfer of command skills training into the workplace and a behavioural research tool.

2.2 Introduction

The performance of incident commanders in the UK FRS is influenced by many factors, which are summarised in Figure 5. These factors include incident characteristics, individual characteristics and the nature of the command team (Butler et al., 2020). The role of incident commander clearly involves a broad range of non-technical skills, which have been highlighted by national bodies. For example, The Department for Communities and Local Government (DCLG, 2013, p. 32) published a health and safety framework specifically for the operational environment of the UK FRS which stated that: “Fire and Rescue Authorities should consider the impact of ‘human factors’ on the safe, effective and timely resolution of an incident. This includes…human and individual characteristics that influence the behaviour of teams and individuals. Understanding these ‘human factors’ is critical to effective health, safety and welfare management.” Guidance published by the National Operational Guidance Programme (N.O.G., 2021g) identifies command skills, which overlap with those from other safety–critical settings and have been linked to the high-profile incidents involving the UK FRS. In the same year this guidance was published, Butler et al. (2020) had conducted a systematic assessment of the non-technical skills that underpin effective incident command and produced a prototype skill set (see Section 2.4.1). However, there was no means to assess the performance of these skills by incident commanders in the UK FRS. Behavioural marker systems represent one means of doing so, and they have been developed in many other high reliability industries from aviation to healthcare (Flin et al., 2008).

The development of a behavioural marker system would enable the National Fire Chiefs Council (who are the professional voice of the fire and rescue service) to meet some of the challenges that are faced in relation to incident command, including improving the selection, assessment and development of incident commanders and improving their decision making in multi-agency environments when under pressure and dealing with complex and major incidents (Chief Fire Officers Association, 2015). Study 1 sought to refine the existing prototype set of command skills and develop a behavioural marker system to facilitate their evaluation. Finally, a preliminary evaluation of the system was conducted.

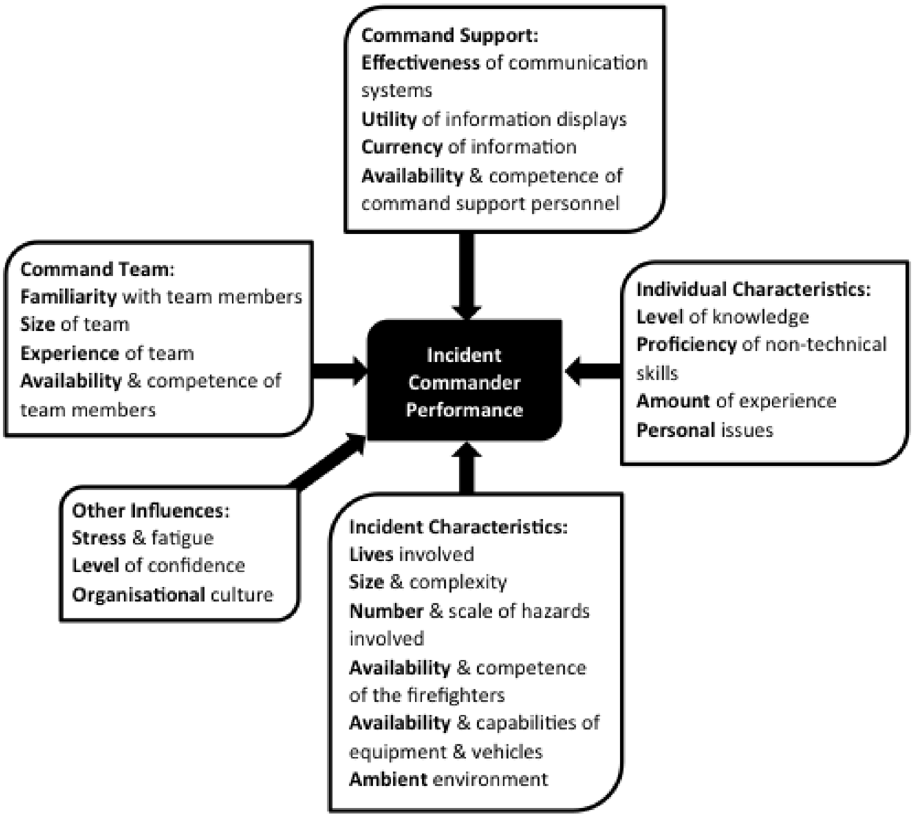


Figure 5: *Factors That Influence the Performance of Incident Commanders (adapted from Youngson, 2016).*

High-reliability industries distinguish between a person's technical and non-technical skills: Technical skills include knowledge of procedures, hazards and equipment, whereas non-technical skills include decision making, interpersonal communication and personal resilience. Historically, the UK FRS has emphasised the importance of technical skills. However, incident commanders within the FRS have a multi-faceted role, which requires them to have both technical and non-technical (command) skills (N.O.G., 2021d). The command skills impact on their ability to work in challenging environments, which are often dangerous and dynamic as well as being emotionally charged and stressful (HSE, 2010a). The role combines decision making with a variety of other non-technical skills (Klein, 2008).

The UK FRS has experienced a number of tragic, high profile incidents involving the deaths of members of the public and/or firefighters (see Section 1.1). Investigations of these incidents have implicated deficiencies in non-technical skills such as situational awareness, communication, teamwork, decision making, and leadership (Torrie, 2012; Greater Manchester FRS, 2016, pp. 4-5; Watterson, 2015; Grenfell Tower Inquiry, 2019). Such deficiencies in non-technical skills mirror those revealed by accident investigations and/or public inquiries into other high-reliability industries (e.g., aviation, marine, oil and gas exploration and healthcare; see Weick, 1990; Air Accident Investigation Bureau, 1990; Marine Casualties Investigative Body, 2012; Crichton et al., 2005; and Health Commission, 2009). Further, investigation of aircraft accidents, in particular, confirmed the need to provide training that would support the development of non-technical skills (e.g., Flin et al., 2008), which became known as crew resource management (CRM) training (Helmreich et al., 1999). The efficacy of this training is validated using a behavioural marker system that consists of a set of observable domain-specific behaviours that relate to the set of non-technical skills (Klampfer et al., 2001) together with a rating system for those skills (Flin &

Martin, 2001). This approach has also been adopted for surgeons, anaesthetists, ships officers and nuclear power reactor engineers (see Devitt & Holford, 2010; Fletcher et al., 2004; O'Connor, O'Dea, Flin, & Belton, 2008; Yule, Flin, Paterson-Brown, Maran, & Rowley, 2006).

2.3 Method

2.3.1 Study Design

The primary design requirements for the THINCS system were for it to be used whilst an incident commander was in charge of an incident or training exercise and to assess the command skills of individual UK FRS tactical incident commanders. National Operational Guidance (N.O.G., 2021f) details the four levels of UK FRS command: initial, intermediate, advanced, and strategic. Of these, the initial, intermediate and advanced levels relate to tactical command therefore the THINCS system had to be suitable to assess incident commanders from these levels.

A behavioural marker system is comprised of two main components, i.e. a set of non-technical skills plus associated behavioural makers and a rating scale. Butler (2016) had identified a prototype set of command skills for UK FRS incident commanders in accordance with a set of design criteria adapted from those used to develop the NOTECHS non-technical skills for European airline pilots (Flin et al., 2003). These criteria meant that each command skill and sub skill had to be:

- As discrete from each other as possible given the interdependence of the skills.
- Clearly defined using UK FRS terminology.

- The minimum number of critical command skills and sub skills as necessary by following the principle of parsimony (Flin et al., 2003).

Butler (2016) identified the command skills using a mixed methods design that presented data gathered about command skills to a group of subject matter experts. These data were gathered from a literature review of command skill-related research, an online survey of UK FRS incident command training managers about their current provision of command skills training, and the themes from a deductive thematic analysis (Braun & Clarke, 2006) of semi-structured interviews[5] of incident commanders exploring their use of command skills. Over two full day workshops, five selected subject matter experts were given a grid exercise and sorting tasks to refine an original set of 59 command skills themes into a prototype set of six command skills (Table 1; Butler, 2016; Butler et al., 2020).

Table 1: *UK FRS Prototype Set of Command Skills (Butler, 2016)*

Skill	Sub-skills
Inter-personal communication	Listening
	Communication style
	Briefing
	Debriefing
	Command transfer
Situational Awareness	Information gathering
	Understanding information
	Anticipates incident development
Teamwork and Interoperability	Consultation
	Cooperation
	Delegation
	Team formation

[5] The interviews formed part of an earlier study investigating the decision making of incident commanders (see Cohen-Hatton et al., 2015).

	People orientated
Personal Resilience	Thinking time Stress and fatigue management Confidence
Assertive, Effective and Safe Leadership	Setting and maintaining standards Values and supports others Leadership style Competence Performance management Safety leadership
Effective Decision Making and Planning	Intuitive decision making Analytical decision making Problem solving Planning

Study 1 was designed to conduct the next phase of behavioural marker system development. That is, to develop a set of behavioural markers for each sub skill before going on to establish a rating scale. Throughout both of these processes, the subject matter experts were free to amend the prototype set of command skills until a final optimal set was created. Finally, once the THINCS system was complete it underwent a preliminary evaluation[6].

The design criteria for the behavioural markers were adapted from the characteristics of a good behavioural marker identified by Klampfer et al. (2001). Each behavioural marker had to: be specific and observable, show a causal relationship to performance outcome, use FRS-specific language that reflects the operational environment, employ simple phraseology, and describe a clear concept. The design requirements for the rating scale

[6] Besides the preliminary evaluation of THINCS being conducted to assess the system's effectiveness it was also done to inform the development of an associated app.

were that it had to be straightforward to use and enable the FRS to determine if the performance of command skills by an incident commander was satisfactory or not.

The approach employed here mirrored those conducted within other high reliability industries, which entailed using workshops involving subject matter experts to identify the behavioural markers (see Crichton & Flin, 2004; Fletcher et al., 2004; Mitchell et al., 2011; O'Connor & Long, 2011; Roberts, Flin, & Cleland, 2015).

2.3.2 Participants

Five subject matter experts were selected based on the following criteria: one representative from each level of command (initial, intermediate, advanced and strategic); all operationally active incident commanders; and either a member of the National Operational Guidance Programme project to revise and develop UK FRS incident command policy, or an incident command training manager, or a member of the National Fire Chiefs Council's National Command and Control Working Group. The use of subject-matter experts allowed access to their domain knowledge, and to characterise patterns of behaviour that are indicative of each non-technical skill (for a review, see Klein, 2008). The five experts had also worked on the earlier research project to develop the prototype of command skills. A larger number of subject matter experts would have been desirable, greater access was not feasible due to the commitment required and the nature of the roles of the more senior subject matter experts. In total they participated in two workshops and a preliminary evaluation although, three were unavailable to participate in the evaluation and were replaced with two further subject matter experts bringing the total to four.

2.3.3 Procedure

2.3.3.1 Subject matter expert workshops. The UK FRS does not pool learning from incidents across the service. Also, the service does not have a national accident investigation organisation such as the aviation and maritime industries that make their independent investigations publicly available. In such circumstances, where data may not exist or is very challenging to collect, subject matter experts are invaluable (Louie & Carley, 2008). Furthermore, the use of subject matter experts to help to shape and refine sets of non-technical skills, and develop behavioural marker systems, is a standard practice (see Flin et al., 2003; Fletcher et al., 2004; O'Connor & Long, 2011). The workshops took place in classrooms or conference rooms at FRS fire stations or training venues.

Throughout the series of workshops and the preliminary evaluation, the prototype set of command skills and their sub skills was kept under review to ensure that it followed the principle of parsimony and retained the minimum number of skills necessary. This was one of the main design criteria used to develop non-technical skills for pilots (Flin et al., 2003). The subject matter experts were free to amend the prototype set of command skills based on their knowledge and experience and use of the system. This process removed or reconfigured some of the sub skills.

The first of the workshops involved the subject matter experts discussing and identifying specific, observable behaviours indicative of good and poor performance for the sub skills of each command skill. The identified behaviours were reviewed and refined through further discussion into a set of exemplary behaviours for each sub skill. These behaviours were described as action statements that either directly or indirectly denoted the presence of a skill (e.g., an order indicating a type of decision making; Cohen-Hatton et al., 2015; Cohen-Hatton & Honey 2015; Fletcher et al., 2004). The resulting behavioural makers were further refined by cross-checking its content with UK FRS national occupational standards (Skills for Justice, 2013a; 2013b; 2014) and guidance for incident command to ensure they were comprehensive.

The workshops also developed a rating scale. A five point, quality-based Likert scale was selected, where a score of 4 (good) represented consistently high performance of the skill and 0 (unobserved) where a skill should have been demonstrated but was omitted, and so represented a significant risk to others. An additional rating of 'not observed' was incorporated to record when a skill was not relevant to the situation under observation. The subject matter experts agreed that an individual's command skill rating would be the mean of its sub skill ratings (excluding 'not observed' ratings). The workshops also created template documents for the recording and feedback of performance to incident commanders and the behavioural marker system was named THINCS.

2.3.3.2 Preliminary evaluation. Four subject matter experts were trained to score an incident commander's command skills performance. One day of training was provided on how to use the behavioural marker system. On two further days the four Raters independently scored the command skills performance of six incident commanders using THINCS. Two of the Raters had been involved in the development of the system and 2 were naïve to the system. The recordings were a mixture of incident commanders responding to real incidents (IC 1, IC 2 and IC 6), live exercises (IC 3 and IC 4), and a computer simulated exercise (IC 5). Each Rater completed a questionnaire at the conclusion of the evaluation. The questionnaire focused on the system's comprehensiveness, observability and utility. The Raters had to either rate their views on a scale of 1 – 5 (1 = very difficult to 5 = very easy) or respond to series of questions with binary answers to represent their point of view (see Appendix 1).

2.4 Results

2.4.1 Subject Matter Expert Workshops

The workshops refined the existing prototype set of command skills and sub skills into a final skill set (see Table 2). For example, under the ‘interpersonal communication’ command skill, the ‘briefing’, ‘debriefing’ and ‘command transfer’ sub skills were merged into one ‘briefing’ sub skill and its definition accordingly amended. Likewise, within the ‘teamwork and interoperability’ command skill, the ‘delegation’ and ‘team formation’ sub skills were merged into ‘team formation’. Exemplary behaviours for the sub skills of each command skill that formed the THINCS system were identified by the workshops. For example, Table 3 demonstrates the good and poor practice behavioural markers for the sub skill ‘thinking time’ of the ‘personal resilience’ command skill. Table 4 depicts the rating scale developed by the subject matter experts and used during the preliminary evaluation. Figures 6 and 7 show the final versions of the documents upon which behavioural observations were made and feedback provided.

Table 2: *Final Set of UK FRS Command Skills*

Command Skill	Sub Skills
Assertive, effective & safe leadership	Setting & maintaining standards of performance Values & supports others Leadership style Competence[7] Safety leadership
Effective decision making & planning	Intuitive decision making Analytical decision making Planning
Interpersonal communication	Listening Communication style Briefing
Personal Resilience	Thinking time Stress & fatigue management Confidence
Situational awareness	Information gathering Understanding information Anticipating incident developments
Teamwork & interoperability	Cooperation Team formation People oriented

[7] Competence with respect to leadership in the THINCS system does not refer to the incident commander's competence (for this see Chapter 1). Instead, it refers to how the incident commander uses their competence to influence their command, e.g., employs their technical knowledge to support their point of view in a meeting.

Table 3: *Exemplary Behavioural Markers for the Sub Skill 'Thinking Time' from the 'Personal Resilience' Command Skill of the THINCS System*

Personal Resilience Sub Skill: Thinking Time

Behavioral Markers

Good Practice	**Poor Practice**
Requests 'quiet' at the command location	Fails to control the distractions at the command location
Creates a barrier when necessary to distractions, or physically removes self away from them	Fails to control others wishing to communicate, jeopardising the successful transfer of critical information
Allocates appropriate command tasks to others to create time to think	Becomes overloaded as fails to delegate roles and work appropriately
Maintains appropriate spans of control to effectively manage workload	Becomes distracted and fails to respond to critical information and act in a timely manner

Table 4: *The Rating Scale of the THINCS Behavioural Marker System*

Rating		**Description**
4	Good	Performance was consistently high/exemplary
3	Satisfactory	Performance was acceptable but with room for improvement
2	Marginal	Performance was inconsistent and unacceptable with improvement required in specific areas
1	Poor	Performance was cause for concern with substantial improvement required
0	Unobserved	Omission to perform skill may have endangered the lives of firefighters and/or dangerously allowed the incident to escalate; serious remediation required
NO	Not Observed	Skill/sub skill not observed as it did not apply to the context

UK FRS THINCS Behavioural Marker System v1.0 — Observation Pro Forma v1

Rater:	Date:	
Incident Commander:	Start Time:	Finish Time:

Item Number	Time	Command Skill & Sub Skill Number	Incident Commander Activity	Behavioural Observations
		Assertive, effective & safe leadership 1 2 3 4 5; Effective decision making & planning 1 2 3; Interpersonal communication 1 2 3; Personal resilience 1 2 3; Situational awareness 1 2 3; Teamwork & interoperability 1 2 3		
		Assertive, effective & safe leadership 1 2 3 4 5; Effective decision making & planning 1 2 3; Interpersonal communication 1 2 3; Personal resilience 1 2 3; Situational awareness 1 2 3; Teamwork & interoperability 1 2 3		
		Assertive, effective & safe leadership 1 2 3 4 5; Effective decision making & planning 1 2 3; Interpersonal communication 1 2 3; Personal resilience 1 2 3; Situational awareness 1 2 3; Teamwork & interoperability 1 2 3		
		Assertive, effective & safe leadership 1 2 3 4 5; Effective decision making & planning 1 2 3; Interpersonal communication 1 2 3; Personal resilience 1 2 3; Situational awareness 1 2 3; Teamwork & interoperability 1 2 3		

Figure 6: *The THINCS Observation Pro Forma*

Ratings and Feedback

Command Skill		Sub Skill	Key Observations for Feedback	Sub Skill Rating						Command Skill Rating
Assertive, effective and safe leadership	1	Setting & Maintaining Standards of Performance		4	3	2	1	0	NO	
	2	Values & Supports Others		4	3	2	1	0	NO	
	3	Leadership Style		4	3	2	1	0	NO	
	4	Competence		4	3	2	1	0	NO	
	5	Safety Leadership		4	3	2	1	0	NO	

Figure 7: *THINCS Feedback Pro Forma Showing the 'Assertive, Effective and Safe Leadership' Section*

2.4.2 Preliminary evaluation

Figure 8 shows the mean rating for each command skill awarded by the Raters and illustrates that the system identified consistent performance differences between the commanders. Thus, one incident commander (IC 2) had higher ratings across the command skills than the remaining incident commanders. Two incident commanders (IC 1 and IC 3) had consistently lower ratings than the remaining incident commanders. ANOVA with the within-subjects variables of incident commander (IC1-IC6) and command skill as factors revealed an effect of incident commander ($F(5, 15) = 6.75$, $p < .005$, $\eta p^2 = .69$), no effect of command skill ($F < 1$), and no interaction between these factors ($F(25, 75) = 1.12$, $p > .34$, $\eta p^2 = .27$); and post hoc tests with a Bonferroni correction revealed a difference between IC1 and IC2 ($p < .05$). These results suggest that THINCS is a sensitive tool, insofar as it is able to distinguish between good and poor behaviours in relation to different sub-skills (Flin et al., 2008)

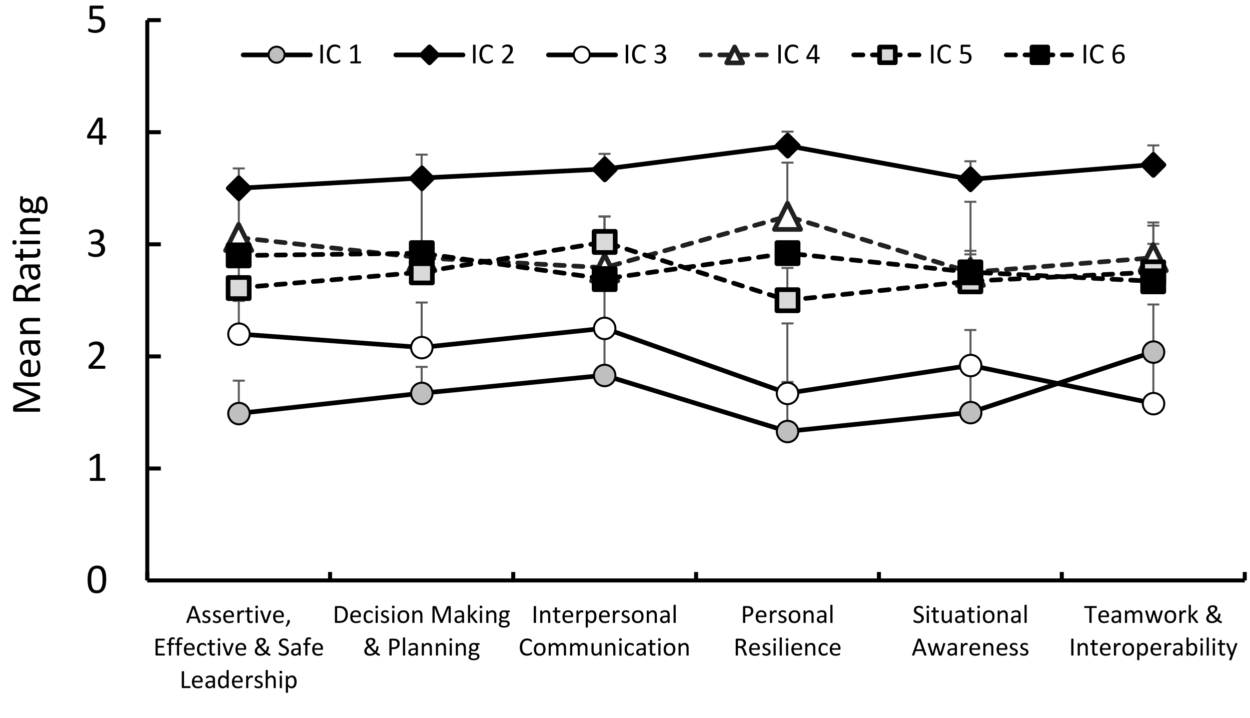


Figure 8: *Mean Ratings (+SEM) for Each Non-technical Command Skill for Six Incident Commanders (IC1-IC6).*

An analysis of inter-rater reliability on the mean scores pooled across all command skills, which were rounded to the closest whole numbers revealed a Cronbach's $\alpha = 0.87$. Cronbach's alpha was selected as it is regarded as the most common measure of scale reliability (Field, 2009).

The Raters' questionnaire revealed that they all considered the system to be comprehensive (100% positive responses). Overall, on a scale of 1 to 5 with 1 being very difficult and 5 very easy, the Raters found it was straightforward to interpret the skills from the observed behaviours (mean rating: 3.75) and, using the same scale, found the documents relatively easy to use (mean rating: 3.25).

2.5 Discussion

The UK FRS faces significant challenges in the future in relation to incident command (Chief Fire Officers Association, 2015). These challenges include incident commander decision making, the reduction in opportunities to gain experience of incident command, and a need to improve the selection, assessment and development of incident commanders. The THINCS system was designed to support the UK FRS in meeting these challenges by enabling the critical command skills to be rated in a transparent way that allows ready feedback to incident commanders after assessments, training exercises and incidents. Several design challenges had to be overcome when developing the system. These included designing the system to be as simple to use as possible, and that the command skills and sub skills it embodied were meaningful to incident commanders, observable and discrete. Feedback from the Raters who participated in the preliminary evaluation indicated that the system contained only those skills and sub skills that were relevant for safe and efficient incident command. Skills, that if performed well, would be associated with good outcomes, but if absent or poorly performed may endanger firefighters and others and/or cause an incident to dangerously escalate.

It was important to develop a credible system with high utility to encourage the future uptake of THINCS. To support its credibility, the system was designed and developed with incident commanders regarded as subject matter experts. The use of such experts from each level of command was vitally important in identifying the behavioural markers for good and poor practice associated with the system's command skills. Likewise, the system was also informed by current UK FRS incident command guidance and national occupational standards. This approach helped to ensure that the research represented different perspectives from the target population and enabled it to act as a catalyst for change (Guba & Lincoln, 1994). Utility also plays an important role in whether a new tool is taken forward (Fletcher et al., 2004). Feedback from the Raters indicated that the paper-based documents used to record their observations and to provide feedback were adequate but not exemplary. This led to the system being implemented as a mobile app that affords greater ease of use in a range of operational, training and research settings. The app calculates the command skill rating and collates time-stamped observations about each sub skill, which can be scored and used to provide immediate feedback to an incident commander. It also produces data that can be transferred and used to form the basis of an organisational database that may be analysed to inform training and future practice. (see appendix 2).

The principal limitation of the initial research conducted to develop the THINCS system was that it had only been subjected to a preliminary evaluation. A larger-scale evaluation would be needed to more fully explore its reliability by assessing: the stability or consistency of ratings of the same incident commander made over time made by a given

Rater/s; the consistency with which different Raters score the same incident commander/s (i.e., the inter-rater reliability); and the relationships (i.e., inter-correlations) between the different sub-skills (i.e., internal reliability; see Bryman, 2012). However, the preliminary analysis of inter-rater reliability suggested that a set of independent Raters provided similar overall rankings of the performance of six incident commanders. One feature of this preliminary analysis was the rankings for a given incident commander were similar across the different command skills. This might either reflect the fact that the Raters' judgements about the different command skills were not independent, or that the different command skills are inter-dependent.

2.6 Conclusion

The THINCS system complements those core non-technical skills described by Flin et al. (2008) and echoes the command skills within the UK FRS national guidance for incident command (N.O.G. 2021d). However, the system is domain specific: it was framed in UK FRS terminology and developed with the assistance of incident commanders regarded as subject matter experts from within the service and was based on a prototype of command skills similarly developed. The THINCS system provides the UK FRS with a tool to prepare its incident commanders more effectively and to monitor and determine the influence that command skills have on incident outcomes. However, further, more extensive evaluation of the system was needed to confirm its validity, reliability, sensitivity and usability as an assessment tool aimed at improving the performance of command skills to reduce human error and contribute to the safety of UK FRS operations. This further evaluation formed part of studies 2a, 2b and 3b.

Chapter 3

Studies 2a & 2b: The THINCS Behavioural Marker System Evaluation

3.1 Abstract

The THINCS behavioural marker system was developed with the support of the NFCC to provide UK FRSs with a means to assess incident commander performance of non-technical, command skills (Butler et al., 2020). It was also envisaged that THINCS could be used in a research context. To more fully evaluate THINCS, two further studies were conducted. The principal aims of Study 2a was to evaluate the reliability, validity, accuracy and usability of THINCS, and also to investigate any underlying relationships between the six command skills. Additionally, the usability of the THINCS app was also evaluated. Study 2a involved seven UK FRSs that permitted personnel to be trained as THINCS Raters to assess incident commander volunteers from within their services. At the conclusion of the study, a series of focus groups with Raters from each FRS focused on the validity and usability of the system. The results of Study 2a demonstrated that the THINCS system was reliable, valid and had a good degree of usability as did the THINCS app, which was more efficient and easy to use and navigate. It also showed that the command skills were all highly correlated. Study 2b was used to evaluate the sensitivity of THINCS through a re-analysis of secondary data from earlier research in which groups of incident commanders had received different forms of training before responding to a "live burn" (Cohen-Hatton & Honey, 2015). Study 2b revealed group differences in the mean number of incident commander observations per activity, but not in the quality of the different command skills. The results also provided further evidence that the THINCS command skills were interrelated.

3.2 Introduction

An established challenge for the UK Fire and Rescue Service (FRS) as a high reliability industry is the need to maintain command competence against a reducing number of operational incidents (Chief Fire Officers Association, 2015). Competent incident commanders are essential to safeguarding firefighter safety in accordance with the firefighter's maxim which advocates greater risks should be taken to, for example, save lives (N.O.G., 2021h). A cornerstone of safeguarding them from harm is the preparation of UK FRS personnel to make decisions in hazardous, fast-moving environments when under pressure (HSE, 2010a). This involves the use of command skills as well as technical skills by incident commanders (N.O.G.2021d)

There is a clear link between the practice of command or non-technical skills and safety outcomes within high reliability organisations and industries. Accidents are more likely to occur when the practice of command skills is poor compared to when they are performed well (Flin et al., 2008). Over the last 40 years high reliability organisations and industries have responded to this relationship by developing training programmes based on research-based bespoke sets of non-technical skills, for example, crew resource management training within the aviation industry (Thomas, 2018). Traditionally, such industries have developed behavioural marker systems to evaluate non-technical skills training and measure their performance (see Devitt & Holford, 2010; Fletcher et al., 2004; O'Connor et al, 2008; Yule et al, 2006). Klampfer et al. (2001) defined behavioural marker systems as frameworks comprising a set of safety and efficiency-related non-technical skills broken down into some constituent parts (e.g., sub skills) highlighted by a set of exemplary behaviours for good or poor performance.

Behavioural marker systems are psychometric tests. That is, tests that apply numbers to latent variables of human performance that cannot be directly measured, for example, decision making and situational awareness. Non-technical skills, such as the command skills of the THINCS system, are constructs (i.e., observable entities that can only be inferred from overt behaviours; Raykov & Marcoulides, 2011). However, when referring to constructs it is important to acknowledge the behaviours that represent them do so only partially (Raykov & Marcoulides, 2011). Consequently, behavioural markers do not represent the whole construct of, for example, decision making. Yet, whilst behavioural markers are only indicative of a construct, it is this relationship that enables a construct to be inferred from them when observed (Raykov & Marcoulides, 2011). Therefore, the definition and structure of these constructs are important with respect to the measurement characteristics validity, reliability, and sensitivity, where sensitivity is a measure of the ability of the system to distinguish between good and poor behaviours in relation to each sub-skill (Flin et al., 2008).

THINCS was developed with the support of the NFCC (with additional funding from the ESRC) to enhance the development and assessment of incident commanders and improve incident ground safety by reducing human error (Butler et al., 2020). Given the links to safety, it is essential to evaluate THINCS with respect to its measurement characteristics (Thomas, 2018). Flin, O'Connor and Crichton (2008) identify the key properties of behavioural marker systems: validity, reliability, sensitivity and usability. Flin et al. (2008) argue that behavioural marker systems with these psychometric properties will have improved measurement quality. More generally, Holt, Boehm-Davis and Beaubien (2001) specify an interdependent set of critical design steps that flow from construct definition to measurement design to objective confirmation of the validity, reliability, and sensitivity of the measures. A focus of the identification of the command skills that underpin THINCS was their definition as constructs and resultant structure along with a preliminary evaluation of their validity, reliability and sensitivity (Butler et al, 2020).

Validity can be defined in terms of whether a test assesses the construct/s that it was designed to assess, whereas reliability can be defined in terms of the consistency of the measurements returned by that test (stability, inter-rater reliability, and internal reliability; Bryman, 2012). However, Raykov and Marcoulides (2011) highlight that validity is an evolving and continuous process because of the challenges associated with determining that, for example, a behavioural marker system like THINCS is precisely measuring a specific sub skill. Validity is built up from evidence across a range of tests that indicate the degree to which the workplace performance can be interpreted from the results of a given test.

There are several ways to gauge the validity of a test. Face validity refers to the relevance and representativeness of a test and how well it is perceived by practitioners to measure its constructs (Urbina, 2004). Whilst it is not an actual psychometric measure of validity, it is important to achieve a high degree of face validity for the purposes of practitioner acceptance of the test and its results with respect to future uptake (Fletcher et al., 2003). Face validity is often determined by consulting subject matter experts (Bryman, 2012). In contrast, content validity refers to the degree to which the elements of a test adequately assess the constructs it purports to measure (Urbina, 2004). It may be regarded as a conceptual test of an instrument's ability to measure what it was developed to measure; and is also engendered by using subject matter experts, in the design of the test, to determine and define its elements so that they align with its constructs (Raykov & Marcoulides, 2011; Urbina, 2004). Also, content validity can be reinforced by using subject matter experts, independent of the initial design of a test, to judge whether or not the elements of test reflect its constructs (Urbina, 2004). The preliminary evaluation of THINCS established that the system had high content validity based on a very small number of participants, some of whom were involved in the development of the system (Butler et al., 2020). This study sought to replicate that finding using a much larger group of participants who did not engage with its design.

Raykov and Marcoulides (2011) define reliability as the lack of systematic and random error in a psychometric tool. Systematic error is repeatable and consistently occurs but is not related to the construct being measured, and it either positively or negatively biases the measurement resulting from use of the tool. This may occur at an individual level due to a personal characteristic such as dyslexia, or at a group level and reflect a training content error. Consequently, it is vital to minimise systematic error (e.g., by using colour to categorise constructs to assist those with dyslexia) to ensure that a psychometric tool is trusted (Raykov & Marcoulides, 2011). Like systematic error, random error is not related to the construct being measured but may be attributed to happenstance such as the temperature of an assessment environment, or the degree to which an individual feels stressed or fatigued, which then affects an individual's performance. Thus, random errors are not wholly repeatable and only occur at the moment an assessment is conducted. It is important to note that in the case of psychometric tests such as behavioural marker systems, systematic and random error applies not only to the individual being assessed, but also to the Rater (De Vellis, 2005). As both can compromise interpretation of the results of a test, it is vital to ensure the reliability of a behavioural marker system. The key reliability checks for a psychometric test are stability, inter-rater reliability, and internal reliability (Bryman, 2012). Stability is established by administering the same test (e.g., rating the same video of an incident commander) at two separate points in time and then correlating their results (Raykov & Marcoulides, 2011). The greater the correlation between the two test results, the more stable the test. Inter-rater reliability is the degree to which the scores of two or more Raters agree. A measure of inter-rater reliability is Cronbach's alpha, which compares the results of different Raters with respect to the same test. Internal reliability tests the level of inter-correlation between items measuring the same or similar construct (Flin et al., 2008). For example, the sub skills of a given command skill.

Evaluation studies typically involve the training of behavioural marker system Raters to assess the performance of individuals by observing short video recordings - typically lasting less than 10 minutes – and complete an evaluation questionnaire (Fletcher et al., 2003; Yule et al, 2006; Mitchell et al., 2011; Rutherford et al., 2015). Butler et al. (2020) conducted a preliminary evaluation of the THINCS system in accordance with the methods adopted by earlier studies. The paper-based version of the system was used by four independent Raters to rate six incident commanders. The findings showed the system was sensitive and able to discriminate between different levels of performance. It also demonstrated a high degree of inter-rater reliability. Usability was measured via an evaluation questionnaire and the Raters recorded that THINCS was comprehensive and straightforward system to use. The preliminary evaluation reported in Study 1 examined inter-rater reliability, but it did not investigate reliability of the THINCS system in terms of its stability over time.

Some non-technical skills are interrelated. For example, relationships have been demonstrated between leadership and communication (Edmondson, 2003), communication and situational awareness (Raferty, Stanton, & Walker, 2013), situational awareness and decision making (Moiser & Fischer, 2010), decision making and stress (McLennan, Strickland, Omodei and Suss, 2014) and stress and teamwork (Driskell, Salas, & Driskell, 2018). Exploring the relationships between the command skills of the THINCS system will determine the degree to which they are interrelated.

To summarise, the main aims of this project were to explore the relationships between the command skills, to further investigate the THINCS system's accuracy, validity, reliability and usability (including that of the THINCS app). The THINCS app (see Appendix 2) was developed after the preliminary evaluation in part to improve the system's usability outdoors, and in inclement weather, as well as its usability via a degree of automation. To achieve these aims, two studies were undertaken. Study 2a formed part of an internship of PCB with the NFCC who recruited seven UK FRSs to participate in a seven-month evaluation project. The aim of the project was to build upon the small preliminary THINCS evaluation by using a large pool of data gathered by many Raters to explore any underlying relationships between the command skills of the system. The Rater assessments were also compared to those of a subject matter expert in the use of THINCS (PCB) to determine their accuracy. The use of a single expert might limit the generality of the conclusions that can be drawn, but was a pragmatic choice based on both availability and expertise. The duration of the project also allowed use of the Raters to test the system's reliability over time.

Whilst the majority of the assessments conducted by the participants involved computer simulations, Study 2a included some that were completed at live exercises, which were more realistic. National operational guidance (2021g) issued by the National Fire Chiefs Council recognises that traditionally the UK FRS has relied heavily upon incident commanders naturally developing their incident command skills and gaining experience through their routine exposure to emergencies. However, recent years have seen a dramatic reduction in the number of emergencies attended by the UK FRS, including a 32% reduction in fires. This compromises a commander's opportunities to develop and gain experience and is recognised as a challenge to the future of incident command (Chief Fire Officers Association, 2015). The importance of providing realistic training has long been recognised as an effective way to transfer learning into the workplace (Burke & Hutchins, 2007; Coultas, Grossman & Salas, 2012). National Operational Guidance (N.O.G., 2021i) acknowledges that realistic incident command training is a vital for complementing experience gained naturally as it introduces real risks to firefighters that commanders are required to manage effectively and safely.

Study 2b focused on the sensitivity of the THINCS system and used footage from an experiment conducted by Cohen-Hatton and Honey (Experiment 3; 2015) in which incident commanders responded to 'live burn' exercises (i.e., those involving real fire and increased risks to firefighters). Before responding to the exercise one group of incident commanders had received "goal-oriented training" and another group received training in the extant, standard operational guidance. Goal-oriented training involved using a short, rapid mental checklist where the commander asked themselves: Why they want to make the decision? What they expect to happen? and Are the benefits proportional to the risks? (N.O.G., 2021j). The controls are designed to focus a commander on evaluating goals, anticipating incident developments, and judging risk following the identification of a course of action. The original research found that goal-oriented training made incident commanders more likely to explicitly consider plans and the situation (Cohen-Hatton & Honey, 2015). In Study 2b, the video footage of the participants was assessed using THINCS to examine its applicability to incident commanders responding to more realistic events, and to examine whether the system was sensitive to the impact of the different forms of training on the performance of command skills.

3.3. Study 2a: Longer-term evaluation of THINCS system

This study required the development of training packages for incident commanders and Raters, and for these training courses to be delivered across UK Fire and Rescue Services. The following four components represent the keys steps in Study 2a.

1. To meet the key property of transparency, THINCS familiarisation training for incident commanders, who would later be subject to assessment, had to be developed so that incident commanders were aware of its content and what constituted good and poor performance of the command skills. This training was designed as e-learning so that it could be accessed and worked through by commanders at times that suited their busy work schedules. Two modules were developed, one covering command skills and the other, the behavioural marker system (available upon request). Every incident commander who was assessed using the behavioural marker system gave consent for their data to be used as part of its evaluation.

2. A course had to be created to train FRS personnel to use the THINCS system to rate the performance of incident commanders. Klampfer et al. (2001) recommended that a Raters training course ought to last between 2-5 consecutive days and consist of between 8-12 people with a half-day follow-up session to gather feedback after the system was being used. Flin et al. (2008) described a Rater training course as needing to include the design and use of the rating system, input on the biases that Raters might experience, the concept of inter-rater reliability, practical training and a formal assessment of Rater competence. A 3-day course was designed for eight people based on these premises and included practise

rating sessions that were used to calibrate the delegates' use of THINCS via plenaries facilitated by PCB at the end of each one to compare and discuss their observations, ratings and feedback (see Appendix 3).

3. The THINCS Rater course had to be delivered to FRS personnel from each of the seven volunteer FRSs. Eight courses were delivered by PCB across the different FRSs. Thereafter, the FRSs decided which version of the system they would use (paper-based or app), set up a database based on that generated by the app, and planned how and when they would use of the system during the evaluation period.

4. Follow-up sessions were arranged towards the end of the project to assess the reliability of THINCS in terms of the stability of the Raters scores over time, and to gather feedback on the system (through focus groups).

3.3.1 Method

3.3.1.1 Participants. 52 incident commanders or incident command trainers were trained to use THINCS. Following this training, four participants were withdrawn from the study after unsuccessfully completing the course and a further eight withdrew from the study. From the remaining 40, 33 participants became active THINCS Raters. Their mean age was approximately 47 (mean = 46.58) with a mean length of service of approximately 21 years (mean = 21. 11). The mean length of time served as an incident commander was just over 14 years (mean = 14.03). 17 (51.52%) of the participants were initial level commanders (Watch Managers or equivalent), 15 (45.45%) were intermediate level commanders (Station Managers), and 1 (3.03%) was an advanced level commander (Group Manager). 72.73% of the participants had a qualification in incident command, ranging from a Leading

Firefighters examination through to a Skills for Justice award in Level 6 Advanced Incident Command. 24.24% had a degree/degree-level qualification, or a postgraduate qualification.

3.3.1.2 Procedure. Over the lifetime of the project the 33 participants who actively used the THINCS system conducted 152 assessments of 81 incident commanders, which generated 2005 observations. 136 (89.47%) of the assessments were conducted of incident commanders practising in a computer simulator, 14 (9.21%) were done at live exercises, and two (1.32%) at real incidents. The mean number of ratings conducted was 4.61, and the mean number of observations per incident was 13.19. These results were captured on a database in each FRS and shared with the researcher at the end of the project. Once these data had been collated into a single database, a data cleansing process ensured they were understandable and accurate. This included the removal of unintelligible data or data from incident commanders for whom there was no consent, ensuring manual command skill calculations were correct, and data from the paper-based version of the system had been accurately transferred onto the FRS database.

The accuracy of the Raters scores following their training course was assessed by comparing them to the reference scores of PCB. The measurement used was the mean absolute differences of the participants' mean command skill ratings from the reference ratings for the skills provided by PCB. Stability was assessed using the Raters assessments of the same video performance at the end of their Raters course, and as part of the follow-up session. These parts of the procedure took place in dedicated training rooms equipped with screens and overhead projectors into which a laptop holding the video recordings was connected. The participants were allowed 60 min to view the video (approximately 10 minutes), make their observations, and review, rate and record their feedback using the paper-based version of THINCS.

The focus groups were conducted after the assessment of THINCS stability. The 44 FRS personnel who participated in the focus groups ranged from the roles of Firefighter to Group Manager. Collectively, they represented initial, intermediate, and advanced level incident commanders and included personnel from training, operations and quality assurance departments. 36 participants were THINCS Raters, five were incident commanders who had undergone a THINCS assessment, whilst another three were observers (i.e., local project managers or administrators). The focus groups were used to gather feedback about the validity and useability of the system. Each focus group began with a recap of the project, an explanation of the researcher's role as a moderator, and an illustration of the ground rules. The questions began with generic ones with respect to assessments *per se* that gradually focused down to key questions in relation to the THINCS system (see Appendix 4). Participants were asked to base their responses on their experience of using the system and from the perspective of their current role (e.g., an incident command trainer, incident commander, or training team manager). Transcripts were generated of participant responses to the key questions about the THINCS system for later analysis. These were analysed using a modified form of Braun and Clarke's (2006) thematic analysis using NVivo 12 software (QSR International, 2018). A line-by-line examination of the transcripts codified references with respect to critical aspects of the THINCS system. These were collated into themes that were progressively refined with respect to validity and usability.

3.3.2 Results

3.3.2.1 Accuracy. The mean absolute difference between the participants' mean command skill ratings (across all skills) and those of PCB was .35 scale points (SEM = 0.06). The fact that the score is low suggests that there is good agreement between the ratings given by participants and those of PCB.

3.3.2.2 Command Skills Relationships. Underlying relationships between the command skills was analysed using principal components analysis (PCA), which is a data reduction technique that identifies interrelationships between a set of variables. The process reduces the set to a smaller number of variables called components. PCA was used to examine if any of the six command skills could be combined into a smaller number of components because some are highly correlated. The PCA was conducted on the command skill ratings of the 33 Raters from their final rating. This rating was selected because the Raters would have the most experience of using the system. Any 'Not Observed' ratings were recorded as missing data and these were excluded from the calculations (in accordance with the THINCS system). The Kaiser-Mayer-Olkin measure verified the sampling adequacy for analysis, KMO = .78, which is well above Kaiser's criterion of .5. The analysis identified one component with an eigenvalue over Kaiser's criterion of 1, which explained 61.22% of the variance. The scree plot supported retaining 1 component, which meant there was no need to rotate the data. This demonstrates that all the command skills are highly interrelated (see Table 5).

Table 5: *Component Matrix for the Command Skills of the THINCS Behavioural Marker System*

Command Skill	Component 1 Loading
Personal resilience	.924
Situation awareness	.909
Effective decision making & planning	.777
Assertive, effective & safe leadership	.699
Teamwork & interoperability	.683
Inter-personal communication	.659

3.3.2.3 Reliability: Stability. The first test was completed at the beginning of the project. It was the formal, practical assessment that each Rater took to demonstrate their competence at the end of their training. The retest took place during the follow-up session at the end of Study 2a. The mean duration of the video recordings was 00:11:33 (range 00:09:20 – 00:12:49). Four Raters were unable to attend their session which resulted in 36 out of the 40 Raters participating in the test-retest stability check, one incident commander withdrew consent for their video to be used during the second test, which resulted in the loss of data from a further seven participants. Finally, one participant did not receive the same video in both tests. The final analysis was conducted on the scores from the remaining 28 participants. The overall mean command skill score, across all command skills, was calculated for each Rater's initial test and retest, and the correlation between these scores was (r = .55, p = .002).

3.3.2.4 Validity and Useability. The mean duration of the focus groups was 01:34:22 seconds (range: 01:04:46 – 02:05:52) and the group sizes ranged from 4-9 participants (μ = 6.29). A random sample of four of the focus groups was analysed. The responses of 22 participants were directly related to validity and useability (e.g., Was the app intuitive to navigate?) and these responses were transcribed and analysed using an adapted version of Braun and Clarke's (2006) thematic analysis, and NVivo software (QSR International, 2018). Initially, 22 themes were codified, which were refined to 20 before being further refined to 6 with respect to validity (2) and usability (4; see Appendix 5[8]). The frequency with which each theme was referenced along with illustrative positive and critical quotations for each one is captured in Table 6.

Table 6: *Themes with Respect to Validity and Usability as Identified from the Transcripts of Focus Group Participant Responses to Key Questions*

Themes	References	Illustrative Quotations
Validity		
Assessment instrument	23	The system is a huge benefit where you have an IC who is constantly underperforming on the incident ground or in assessments…although they believe they know every SOP…THINCS allows you to explain to them the [non-technical] reason why they are failing...that is beyond the performance criteria, knowledge and understanding of [current qualifications]
		I found I would record observations that could fit into a number of sub skills
Command skills	39	I think it's quite complete. I'm not finding things I can't put in a box…the sub skills relate to each other and then to me make sense as a command skill
		Personal resilience I found quite challenging and did not mark it across the assessments I did
Useability		
Methodology	57	The system flowed and it was clear what you were marking against and irrespective of the ratings awarded, they were worthwhile ratings to have and led to good feedback
		The current scale does not have an acceptable middle ground. At the moment it's 'marginal' and 'unacceptable'
THINCS app	35	It simplified the system. The training implied it was going to be a long-winded process to learn the system, but then you used the app and you thought, 'Oh! Is that it?', and you realise that it is then very quick. I would only use the app going forward.

[8] Initial thematic coding available upon request

		Ability to save assessments to the device would be helpful when doing back-to-back assessments
THINCS Paper-based Version	22	I think the [feedback pro forma] rates and I think as a debriefing tool its superb.
		Slow, arduous process to transfer [by hand] your observations and feedback over to the database
Implementation Challenges	54	THINCS itself is a good system but it's just having that time to become compliant with it, competent with it, confident with it. But this is no different for bringing in anything else that is new
		Once you have identified [command skill] development areas how are we actually going to provide that development?

Note Text in [brackets] has been inserted to clarify the context of a quotation and typos have been corrected.

3.3.2.5 Content validity. Raters found the command skills and sub skills of the THINCS system to be well-defined, important, and complementary. As the critical quotation above suggests (Table 6), they found some were easier to observe than others, but they also observed some more than others (e.g., situational awareness and safety leadership). The Raters commented on positive performance-related relationships between some of the sub skills (e.g., listening and briefing with respect to interpersonal communication). They reported that in some cases the set of command skills positively broadened their assessment horizon and got them to judge aspects of performance they had not done so previously. Finally, a small number of participants expressed a desire to see more streamlined set of command skills to simplify the THINCS assessment.

3.3.2.6 Face validity. Information was captured from the Raters with respect to the behavioural markers, performance observation and the perceived benefits of the THINCS system. Overall, they reported that the behavioural markers were helpful and useful when assessing. They recognised that the system was sensitive to non-technical issues that impaired performance but were not currently covered within the UK FRS National

Occupational Standards. They expressed how the THINCS system enabled them to better understand and provide feedback on the performance of incident commanders in support of their development. Overall, THINCS was valued as an assessment tool.

3.3.2.7 Usability. The Raters much preferred to use the THINCS app over the paper-based version of the system. The app simplified the assessment process and was intuitive, easy, and more efficient to use because of its automated features, such as, command skill calculations and access to feedback via the icons. Raters also expressed that compared to the paper-based version, it was straightforward to use outside. The Raters identified a number of positive improvements that could be made to the THINCS app to make it easier to use. For example, for it to store assessments on the tablet, to record spoken observations and feedback, to capture video footage of performance, and to edit observations. However, it was recognised that the paper-based version was always available to use compared to the app which could be affected by mobile phone coverage and connectivity issues. These issues would prevent a completed assessment from being e-mailed to an administrator ahead of completing another assessment resulting in its deletion. The main drawback with the paper-based version was time taken to process the observation and feedback records and transfer them onto the database.

Overall, the Raters were positive about the THINCS methodology: appreciating how it flowed, its similarity to other assessment models, and highlighting its objectivity and potential use on the incident ground. The feedback it produced was appreciated with respect to its succinctness and clarity. Whilst in some cases the Raters were satisfied with the rating scale, in general, they were critical of the 5-point rating scale for what they perceived as its lack of an acceptable mid-point score that represented a marginal yet unacceptable performance. The database did not attract many comments, but there was a view that it’s format could have been simpler.

Second only to the generally favourable comments on the THINCS methodology, were concerns raised by the Raters with respect to the implementation of THINCS assessments into their organisations. Key organisational challenges identified were: The lack of available command skills training to meet any development needs arising out of a THINCS assessment; Raters maintaining their THINCS rating skills; the tension between the behavioural marker system and current National Occupational Standards and assessment practices that do not have a behavioural dimension; and the additional resources required to use THINCS alongside current technical assessment methods.

3.4 Study 2b: Evaluation of THINCS system sensitivity

Study 2b used archival data to conduct a further assessment of THINCS. The original study by Cohen-Hatton and Honey (2015) sought to assess the impact of goal-oriented training on the process of decision making in a live burn exercise. Each participant followed the same sequence of events. Participants received a one-hour training session. One group received standard training, based on national guidance at the time (2014). It comprised of a short exercise which involved watching a video of an emergency that was paused at key decision points so the participants could be asked what action they would take and why. The decision control process (DCP) group received similar training except it included information about the use of decision controls (Why am I doing this? What do I expect to happen? and Are the benefits worth the risks?). Also, when the video was paused in their exercise, and they were asked what actions they would take next, they were guided to answer with reference to the decision controls. Within an hour of undertaking the training all participants then completed a live burn exercise (i.e., a realistic exercise involving fire and its associated hazards), and their command was captured on a helmet-mounted video camera. The basic scenario involved a fire in a 1st floor flat within a 3-storey block with a young child reportedly involved. During the exercise the same number and type of unexpected events were applied to each participant (e.g., reports of a baby screaming on the ground floor or the sudden failure of the water supply). The exercise involved information and characters that an incident commander would expect to receive or interact with during the response to such an emergency. None of the role-players involved were aware of the training provided to the participants. In Study 2b, the footage of the participants responding to the live burn was coded by PCB using THINCS. The coding was conducted blind with respect to group membership.

3.4.1 Method

3.4.1.1 Participants. Videos from 21 incident commanders out of the 24[9] from two UK Fire and Rescue Services (Hampshire and the Isle of Wight) who consented to participate in the study included in Cohen-Hatton and Honey (Experiment 3, 2015) were used for this study. Their informed consent included their video footage being used for research purposes. The sample included Level 1 commanders, who would initially be in charge of an emergency (n = 9: DCP = 5, Standard = 4), and Level 2 commanders, who would take over from the Level 1 commanders when necessary (e.g., due to the involvement of higher risks or levels of complexity (*n* = 12: DCP = 6, Standard = 6). A questionnaire in the original study gathered data on the participant's service in the UK FRS, including their years of incident command experience. The mean command experience for the sample was 10.61 years (SEM

[9] The remaining 3 video recordings were unavailable due to the impact of the Coronavirus.

= 0.94, range: 3.00-17.00). For the Standard group it was 10.66 years (SEM = 1.41, range: 3.00-15.30) and for the DCP group it was 10.56 years (SEM = 1.32, range: 3.30-17.00). For Level 1 commanders it was 7.91 years (SEM = 1.37, range: 3.30-14.90) and for Level 2 commanders it was 12.63 years (SEM = 0.95, range: 5.10-17.00).

3.4.1.2 Materials.

3.4.1.2.1 Equipment. The video footage was viewed via a laptop computer (Dell Latitude 5400) on a 14" screen and listened to via earbuds. The videos were in the MP4 format. Version 1.0 of the THINCS app was used to rate the performance of the participants via a 7" Samsung Galaxy Tab A tablet.

3.4.1.2.2 Cohen-Hatton and Honey (2015) Study

The original study initially coded the video footage of the participants based on the primary phases of decision making: Situation assessment (SA), plan formulation (PF) and plan execution (PE). The codes provided a sequence of state transitions from one phase to another. The coding also included evidence of Level 3 situational awareness (anticipation) at the transitions from situation assessment. These data were analysed using lag sequential analysis that characterised how categories of activity were sequenced over the course of the video. Ultimately, the mean conditional probabilities of one phase following another were calculated, specifically from SA to PF or PE, from PF to PE or SA, and from PE to either SA or PF (for further details, see Cohen-Hatton & Honey, 2015). The results were pooled across whether the participants were either Level 1 or Level 2. Table 7 briefly summarises the results (for further details, see Cohen-Hatton & Honey, 2015).

Table 7: *Summary of Findings from Cohen-Hatton and Honey (2015)*

Analysis	Findings
Lag sequential analysis	• The SA-PE transitional probabilities significantly differed between training groups with it happening more likely than would be expected by chance in the Standard group, but not in the DCP group • The PF-PE transitional probabilities for both groups was significantly more likely to happen than would be expected by chance • There was no significant difference between the groups for the PE-SA transitional probabilities
Situational awareness	• Significantly more evidence of Level 3 situational awareness was demonstrated at the SA transition for the DCP group than the Standard group
Manipulation checks	• The DCP group was more likely than would be expected by chance to mention goals, expected consequences and benefits/risks than the Standard group • Time taken to respond unexpected events did not differ between the groups

Note: SA (Situational awareness); PE (Plan Execution); DCP (Decision Controls Group).

These findings suggest that participants who received the goal-oriented (i.e., DCP) training were more likely to give explicit consideration to goals and anticipated consequences, and to show increased anticipatory situation awareness. These findings overlap with the decision making and planning and situational awareness command skills of the THINCS system.

3.4.1.3 Procedure. Each participant's video was rated by PCB using the THINCS methodology. This involves the processes of observation, review and rating. During the observation process PCB noted the activity taking place and recorded the behavioural observations against a sub skill of the relevant command skill (observations can be made against more than one command skill for each activity). They also indicated whether the behaviour represented good or poor practice. PCB was free to pause the video to record their observations. The review process involves a holistic review of the observations made against each sub skill in the context of the activities. Finally, the incident commander's performance is rated: 4 equates to a 'Good' performance, 3 is 'Satisfactory', 2 is 'Marginal', 1 is 'Poor', 0 indicates that the omission to perform a sub skill when it ought to have been resulted in, or could have resulted in, serious consequences. A further rating of 'NO' could also be applied where a skill was not observed. The rating for a given command skill was calculated as the mean of its sub-skill ratings.

The THINCS app automatically produced a worksheet capturing the incident commander's activities and the Rater's behavioural observations, ratings, and feedback. At the end of an assessment the researcher e-mailed the worksheet for inclusion on a database. This had to be done as the THINCS app does not feature the ability to store any data. Once the participant videos had been assessed the data generated was analysed and then compared with the primary findings of the original study.

3.4.2 Results

The mean participant video duration was 00:21:16s (range: 00:09:17s – 00:27:09s). Across the 21 videos, 905 behavioural observations were made against a total of 400 participant activities. The mean number of observations per participant was 43.10 (range: 26-67), the mean number of incident command activities engaged in by participants was 19.05 (range: 10-24), and this resulted in a mean of 2.26 observations per activity (range: 1.44-2.86). In fact, the mean number of observations per activity was significantly higher in group DCP (mean = 2.43, SEM = .15; Level 1 = 2.48, SEM = .20, and Level 2 = 2.40, SEM = .11) than group Standard (mean = 2.08, SEM = .11; Level 1 = 2.10, SEM = .24, and Level 2 = 2.07, SEM = .11). ANOVA confirmed that there was a main effect of group, $F(1, 17) = 5.02$, $p < .05$, $np^2 = .23$, no effect of level, $F(1, 17) = 0.10$, $p = .75$, $np^{2.} = .006$; and no interaction between these factors, $F(1, 17) = 0.02$, $p = .90$, $np^2 = .001$.

Table 8 shows the distribution of observations across the command skills for each training group. This highlights the greatest number of observations for the standard group related to leadership, whilst for the DCP group it was situational awareness.

Table 8: *Distribution of Observations Across the Command Skills for Each Training Group*

	Command Skill					
Training Group	**Assertive, effective & safe leadership**	**Effective decision making & planning**	**Interpersonal communication**	**Personal resilience**	**Situational awareness**	**Teamwork & interoperability**
DCP	106	107	87	33	113	36
Standard	103	92	71	26	99	32
Totals	209	199	158	59	212	68
% of All Observations	23.09%	21.99%	17.46%	6.53%	23.43%	7.51%

The THINCS assessments resulted in a total of 420 performance ratings across all sub skills with the greatest number relating to the ‘satisfactory’ rating (168) followed by ‘not observed’ (96), ‘marginal’ (95), ‘poor’ (52), ‘unobserved’ (7), and ‘good’ (2). The lowest rated sub skill overall was ‘Briefing’ with a mean rating of 1.62 (DCP group = 1.55 and Standard group = 1.70) and the highest ‘People Oriented’ with a mean rating of 2.80 (DCP group = 2.60 and Standard group = 3.00).

The mean scores for the THINCS analysis for groups DCP and Standard are shown in Table 9. Inspection of this table reveals that there was little indication that the command skill scores differed either between the two groups (DCP or Standard) or between the command level (1 or 2). ANOVA with group, command level, and skill as the factors revealed no effect of group, $F(1, 17) = .121$, $p = .73$, $np^2 = .007$, no effect of command level, $F(1, 17) = 0.00$, $p = .99$, $np^2 = .000$, no effect of skill, $F(1, 17) = 2.07$, $p = .07$, $np^2 = .109$, and no interactions between these factors, largest $F(5, 85) = 1.06$, $p = .38$. $np^2 = .059$.

Table 9: *Means (+SEM) for Study 2b*

	DCP		Standard	
Command Level	**1 (*n*=5)**	**2 (*n*=6)**	**1 (*n*=4)**	**2 (*n*=6)**
Command Skill				
Assertive, Effective & Safe leadership	2.27 (.24)	2.33 (.32)	2.35 (.29)	2.33 (.24)
Decision making and planning	2.46 (.29).	2.25 (.29)	1.80 (.38)	2.27 (.28)
Interpersonal Communications	2.00 (.21)	2.27 (.28)	2.13 (.34)	2.14 (.29)
Personal Resilience	2.50 (.26)	2.36 (.24)	2.50 (.22)	2.18 (.36)
Situation Awareness	2.50 (.45)	2.47 (.18)	2.50 (.18)	2.58 (.20)
Teamwork & Interoperability	2.36 (.25)	2.36 (.22)	2.31 (.22)	2.30 (.22)

A principal components analysis (PCA) was conducted on the command skill ratings of the 21 participants. Any 'Not Observed' ratings were recorded as missing data and these were excluded from the calculations (in accordance with the THINCS system). The Kaiser-Mayer-Olkin measure verified the sampling adequacy for analysis, KMO = .81, which is above Kaiser's threshold criterion of .5. Also, all the correlations between command skills were between the acceptable range (i.e., $r \leq .8$ and $\geq .3$) advocated by Field (2009). The analysis identified one component with an eigenvalue over Kaiser's criterion of 1, which explained 62.13% of the variance, upon which all command skills loaded (see Table 10). A scree plot supported retaining 1 component, which meant there was no need to rotate the data. This finding replicates the PCA from the first study in showing that the command skills are highly interrelated with one another.

Table 10: *Component Matrix for the THINCS Command Skills*

Command Skill	**Component 1 Loading**
Assertive, effective & safe leadership	.850
Situation awareness	.847
Teamwork & interoperability	.817
Inter-personal communication	.785
Effective decision making & planning	.780
Personal resilience	.629

3.5 Discussion

As a psychometric test, the THINCS behavioural marker system measures latent variables of human performance (i.e., the command skills): skills only indirectly measured by observable behaviours that indicate the extent to which they are being performed (Raykov & Marcoulides, 2011). To minimise the effects of random error and ensure the objectivity of the THINCS system several techniques were adopted during its design and development. For example, keeping the system simple, limiting THINCS Raters to only assess observed behaviours, and reducing reliance on memory when assessing. Yet, it was essential to test the system's reliability, which Nunnelly (1967) defines as the lack of random error in a psychometric tool and equally, the system's sensitivity and validity (Holt et al., 2001) to establish its bona fides as an assessment tool.

The THINCS behavioural marker system was designed to describe and assess the main command skills that are essential for the safe practice of incident command. As and assessment tool the most important aspect is its validity (i.e., that it measures what it was developed to measure). Study 1 used a series of focus groups to explore the system's validity. Overall, the evidence supported the findings of the preliminary evaluation of the THINCS system and demonstrated acceptable face and content validity. Reliability estimates for the THINCS behavioural marker system began with investigating its inter-rater reliability in the preliminary evaluation that found that the ratings awarded by THINCS Raters highly intercorrelated ($\alpha = .89$). Study 2a built upon this by assessing stability, and found the THINCS system to exhibit stability (i.e., $r = .55$, $p = .002$). This level of stability represents a medium to large statistical effect (Field, 2009), but might be regarded modest in terms of the intended use of the system. However, this stability is evident across an intervening background of experiences (including the use of the system) that were relatively unconstrained.

Usability, like face validity, is an important measure for practitioner uptake. This is achieved by ensuring that a tool is easily understood by users and that it is also easy to use. The feedback from the focus groups suggested that the THINCS app was preferrable over the paper-based version due to, for example, its automatic features and ease of use. Overall, the THINCS methodology was received well, but concerns were raised with respect to the rating scale not having a mid-point rating that represented an 'acceptable' level of performance. The rating scales used within existing systems of other industries vary from four to five-point scales (see Crichton, Moffat, & Crichton, 2017; Flin, Youngson, & Yule, 2016). Traditionally, the central score of 5-point scales has been regarded as a neutral response (cf. Friedman & Amoo, 1999). To reduce the possibility that Raters could make use of such a response, THINCS rating scale was structured to guide the Raters into making a decision between an acceptable or unacceptable performance. In doing so the scale of

‘Good’, ‘Satisfactory’, ‘Marginal’, ‘Poor’ and ‘Unobserved’, does not provide a neutral mid-point (Friedman & Leefer, 1981). Likewise, extreme adjectives were avoided because there is evidence to suggest that Raters avoid them (Pollack, Friedman, & Presby, 1990). A unique feature of the scale is the ‘unobserved’ rating. The subject matter experts justified its inclusion in response to a number of high-profile incidents where incident commanders failed to perform one or more command skills that resulted in harm being caused to members of the public or service reputations (see Torrie, 2012 for a prime example). Sensitivity reflects an assessment instruments’ ability to discriminate between different levels of performance. Whilst the preliminary evaluation found this to be the case, analysis of Rater data generated by the practical assessments of the Rater’s training course for Study 2a found that the mean ratings provided by the participants were similar to those provided by PCB.

Study 2b explored the sensitivity of THINCS by using it to assess the performance of groups of incident commanders given different forms of decision-making training (Cohen-Hatton & Honey, 2015). THINCS did not detect differences in the performance of command skills between these groups. It is worth remembering that in the original analysis provided by Cohen-Hatton and Honey (2015) was quite different from that provided in Study 2b. In Cohen-Hatton and Honey (2015) the process of decision making was coded using the sequences of three simple categories (situation assessment, plan formulation and plan execution). There was no assessment of the quality of decision making processes, or of situation assessment, plan formulation or plan execution. However, THINCS scoring resulted in more observations about activities in group DCP than in group Standard. The basis for this difference is unclear, but it could reflect a general increase in communication between incident commanders with their crews in the DCP group across all types of activity.

The principal components analysis also suggested that the command skills (as evident in their behavioural markers) were highly correlated with one another (see also, Butler et al., 2020).

3.6 Conclusion

The results of Studies 2a and 2b have shown that THINCS and its app are valid and reliable assessment tools, with an ability to discriminate between performances, and a good degree of usability. The fact that THINCS effectively discriminates between different command skills means that it could enhance understanding of how incident commanders perform under a range of important conditions. Study 3a investigated how incident commanders respond to two simulated scenarios. For one scenario, a standard operating procedure (SOP) was available and licenced, whereas for the second scenario departure from a SOP was licenced and the use of operational discretion required. The issue addressed in Study 3a was whether incident commanders used SOPs in a context appropriate manner. In Chapter 5, Study 3b employed THINCS to gain an increased understanding how command skills were used in the two scenarios.

Chapter 4

Study 3a: An Investigation into the Application of Operational Discretion by Incident Commanders

4.1 Abstract

The studies reported in Chapters 4 and 5 sought to better understand how firefighters use rules (i.e., standard operating procedures; SOPs) and deliberative decision making (i.e., operational discretion; operational discretion) at incidents interacts with acute stress. Current operational guidance for UK firefighters combines the provision of SOPs, for routine incidents, with the use of operational discretion, under prescribed conditions (e.g., when there is a risk to human life). However, our understanding of the use of SOPs and operational discretion is limited. Incident commanders (ICs; *n*=43) responded to simulated emergency incidents, which either licensed the use of operational discretion or required use of a SOP. Video footage of IC behaviour during the incidents was used to code their response as involving an SOP or operational discretion, while levels of acute stress were assessed using a blood-based measure and self-report. Incident commanders were *less* likely to use operational discretion selectively in a simulated emergency incident that licensed its use than in one for which use of an SOP was appropriate; an incident commander's level of command did not affect this pattern of results; and the incident that licensed operational discretion resulted in more acute stress than the incident that required use of a SOP. SOPs and operational discretion were not used in the manner prescribed by current operational guidance in simulated emergency incidents. These results suggest that firefighter training in SOPs and operational discretion should be augmented alongside personal resilience training, given its impact on health and wellbeing, but also to improve the deployment of SOPs and operational discretion under stress. Study 3b provides a

more detailed analysis of the (THINCS) command skills used in the two scenarios and investigated further how the approach of incident commanders who used operational discretion in a context appropriate manner differed from those who did not.

4.2 Introduction

The economic impact of fire in the UK in one year alone was estimated to be £8.3B (DCLG, 2011), but fire also has profound environmental and societal impacts. These impacts can be mitigated through the decisions made by first responders (e.g., firefighters). The routine nature of some emergencies means that decision-making can be supported by explicit rules (i.e., standard operating procedures, SOPs) or implicit learnt rules (Epstein, 1994; Gigerenzer, & Goldstein, 1996; Klein, 1993; Sloman, 1996). However, 'unprecedented' emergencies (e.g., Grenfell Tower) require a more flexible, deliberative approach where options are weighed against one another in terms of their potential costs and benefits (Kahneman, 2003; Von Neumann & Morgenstern, 1947). Within the UK fire and rescue service, firefighters have explicit rules (SOPs) for dealing with routine emergencies (e.g., a contained fire in a flat where there was no immediate danger to human life or property), and specified conditions that license departure from them and the use of operational discretion (N.O.G. 2021a). The specified conditions that license such departure include "saving human life, taking decisive action to prevent an incident escalating, and incidents where taking no action may lead others to put themselves in danger". This approach to how decisions are made balances the efficiency of rules with the flexibility afforded by the (conditional) use of deliberation to respond to a wide variety of emergencies. It forms part of the training and accreditation of firefighters, and specifically incident commanders who are responsible for directing the actions of firefighter crews at emergency incidents.

A paradox arises between the conditions under which operational discretion is licensed and converging laboratory research about the conditions that influence the use automatic, rule-based processes rather than deliberative decision making. The conditions where firefighters are licensed to depart from rules and to use operational discretion (e.g., saving human life) is likely to generate acute stress (Lazarus & Folkman, 1984); and acute stress can reduce the capacity for deliberative decision-making, and increase the reliance on rules in a variety of contexts (Porcelli & Delgado, 2009; Schwabe, Tegenthoff, Hoffken, & Wolf, 2012; Starke & Brand, 2012; see also, Janis & Mann, 1977; Kassam, Koslov, & Mendes, 2009; Peters, McEwen, & Friston, 2017; Porcelli & Delgado, 2017). For example, acute (extrinsic) stress exacerbates decision-making biases in gambling tasks, which reflect the operation of automatic processes (Porcelli & Delgado, 2009; cf. Kahneman & Tversky, 1979). Similarly, glucocorticoid and noradrenergic activation results in a shift from goal-directed control of behaviour to automatic, habitual control (Schwabe et al., 2012). Taken together, these results suggest that the very conditions under which firefighter guidance recommends the use of operational discretion rather than the use of an SOP (i.e., when conditions are unprecedented and lives are at risk) might (indirectly) be expected to result in a greater tendency to use an SOP rather than operational discretion.

The paradox outlined above, however, is based upon an extrapolation from laboratory research, where the stressor can be the prospect of the participant (usually an undergraduate student) giving a public talk or having their hands placed in ice-cold water for two minutes (see Porcelli & Delgado, 2009). While these manipulations generate acute stress, they are unrepresentative of the conditions faced by firefighters: who often work in challenging environments, which are characterized as time pressured, with high stakes and involving ill-structured problems (Orasanu & Connolly, 1993). Moreover, the nature of the firefighting role, the decisions that it entails (Klein, 1993), and indeed the individual characteristics of firefighters (Lazarus & Folkman, 1984; for reviews, see Mark & Smith, 2008; Salas, Driskell, & Hughes, 1996), might mean that the results of laboratory research are of little relevance to firefighter decision-making. The field of naturalistic decision making is concerned with just these issues, and studies within this field have revealed important insights into the nature of decision making in the world outside of the laboratory (see Zsambok & Klein, 2014). Our research is in that tradition. However, to the best of our knowledge, no study has assessed either (1) whether firefighters are more (or less) likely to depart from SOPs when the conditions are met to do so, and (2) whether or not those conditions are in fact perceived as stressful by firefighters (Lazarus & Folkman, 1984).

To address these critical gaps in our knowledge, the use of SOPs and operational discretion by incident commanders was examined. Incident commanders in the UK fire and rescue service are multi-faceted: Briefly, they are expected to gather information that is relevant to the incident concerning resources and hazards in order to inform the selection of the appropriate course of action, and to communicate these actions to members of their crews, and other responding agencies. Here, incident commanders responded to two simulated incidents: The Discretion scenario involved a group of children who had fallen into a sinkhole in a remote location, and licensed departure from SOPs on the basis of, for example, saving human life. The Control scenario involved a contained fire in a flat where there was no immediate danger to human life or property, which could be dealt with using the SOP (see Table 11 below). A blood-based assessment of immune system function

(Shelton-Rayner, Macdonald, Chandler, Robertson, & Mian, 2010) and self-report was used to assess the levels of acute stress generated by the two incidents; and cued recall interviews were used to explore the rationales behind the incident commanders' decision making. On the basis of the laboratory research described above, it was predicted that participants would be more reliant on the SOP and less likely to use operational discretion in the Discretion than in the Control scenario; with the Discretion scenario generating higher levels of acute stress than the Control scenario. It was also predicted there would be behavioural differences with respect to the performance of command skills between the incident commanders who used operational discretion appropriately and those who did not. Finally, we examined the potential impact of command level (cf. Klein, 1998; Klein, Calderwood, & MacGregor, 1989) on the use of SOPs and operational discretion, and on acute stress; with the caveat that they were relatively very senior incident commanders.

4.3 Method

4.3.1 Participants

Forty-three incident commanders (42 male) volunteered from 15 UK Fire and Rescue Services (including three of the four UK nations) and provided informed consent for their participation in accordance with local ethical approval through the School of Psychology, Cardiff University. The use of a within-subjects design (with all participants receiving both scenarios) meant that the overall sample size was relatively large, while being determined by the availability and willingness of UK incident commanders to be involved in the research. The participants had a mean length of service of 22.84 years (range: 5.00–40.50 years), a mean length of experience in an incident commander role of 16.38 years (range: 2.67– 30.00 years), and a mean length of service in current role (Level 2 or 3) of 4.05 years (range: .08–24 years). All participants were active incident commanders who were either at Level 2, intermediate (*n*= 32) or Level 3, advanced (*n*=11). Level 2 commanders are command and control middle managers at a tactical level, and Level 3 commanders operate at the tactical level at the scene of large and serious incidents. The participants wore standard issue fire service uniforms during the scenarios. The removal, storage, use and disposal of blood samples, was conducted in accordance with the Human Tissues Act 2004.

4.3.2 Materials

4.3.2.1 Equipment**.** Before undertaking the two simulated scenarios, all participants completed a suite of online questionnaires using Qualtrics software (Qualtrics, 2019). These included a stress-related questionnaire, which combines the Smith Wellbeing questionnaire (SWELL), which focuses on occupational issues (Smith & Smith, 2017), and the Wellbeing Process Questionnaire (WPQ), which focuses on personality characteristics (Williams, Pendlebury, & Smith, 2017). Here, our main interest was in the level of acute stress during the two scenarios; but we also examined whether there was any relationship between chronic stress, as measured in the questionnaire, and our two measures of acute stress (LCC and self-reported stress); as we will show, there was not. The participants also completed a questionnaire to capture details of their operational experience across all levels of command they had practiced.

The simulations were conducted in a purpose-built incident command simulation suite at the Headquarters of Hampshire Fire and Rescue Service. The simulation suite consisted of a series of training rooms and a control room housing the equipment required to control the course of the simulated events: computers, audio and visual monitors, and communications equipment. During a simulation, the control room contained the simulation director, the radio communications role-player, and an XVR-trained technician to manipulate images of the incident. The moving images that represented the scene of the incident were displayed in a training room in which the simulated incident took place. These images were created and generated using XVR software. This room also acted as a holding area for all other role-players.

A large training room (H×L×W: 2.5m×10m×6m) housed the mock command unit and a large monitor used to display a digital film of the changing situation at the scene. GoPro cameras were used to capture activity within this room. A digital clock, placed in the field of view of one of these cameras, enabled key events to be timed. Handheld radios were used for mobilising control centre and incident ground radio communications. A data projector was used to display command support software, such as a decision log and location information, maps and images. Each simulation involved several generic role-players such as command unit officers, police and ambulance officers, along with role-players who were specific to the scenario (e.g., relatives of those involved, an Urban Search and Rescue (USAR) line rescue tactical advisor, Hazard Area Response Team (HART) team leader, and an aerial ladder platform Crew Manager). The command unit officers were trained staff who performed the role at real incidents. They were briefed to support the participants as they would commanders at a real incident and provided with copies of the prescribed radio messages from the mobilising control centre and incident ground.

A smaller room (H×L×W: 2.5m×6m×4m) was used to take blood samples before and after both simulations, and to attach a chest-mounted GoPro camera to capture their conversations and verbalised thoughts. The blood samples were used to provide an assessment of the impact of the two scenarios on a marker of immune system function.

Briefly, leukocytes are white blood cells that are involved in the immune system's first response to threat of ill health caused by foreign bodies or stress. There are different types of white blood cells, with neutrophils representing the majority. One way neutrophils respond to stress is to release reactive oxygen species (ROS) and neutrophils circulating in the blood that have responded to one threat will have a reduced capacity to generate ROS to challenge another. Leukocyte Coping Capacity (LCC) is a measure of the ability of leukocytes (mainly neutrophils) to produce ROS in response to a chemical attack of phorbol myristate acetate (PMA; see Shelton-Rayner et al., 2010). The greater the level of neutrophil reactivity, the greater the ability to cope with stress. This measurement can be considered to represent an individual's level of resilience to stress. For this study, LCC was measured using a test kit produced by Oxford MediStress Ltd (Oxford, UK), which includes a luminometer, heating block, pipette, buffer solution and PMA reagent. For each of the 4 samples per individual, a trained researcher used a disposable blood lancet on a finger to generate a pinprick (10 micro-litres) of blood that was drawn off using a pipette. The blood was transferred to a glass luminometer tube held in a heating block at body temperature (37^{0}C) containing the PMA reagent mixed with a buffer solution. After 10 min the sample was tested by placing the glass tube in a luminometer and a reading of reactivity taken in Relative Light Units (RLU). Lower scores are associated with recent exposure to a stressor, and a reduced potential to cope with future stressors. In fact, the LLC scored were expressed as a ratio: LLC score after the scenario, relative to the sum of this score and the LCC score before the scenario. Using this measure, scores below .50 indicate that the LCC score is lower after the scenario than before it. The LCC scores were complemented by self-report measures during the two scenarios. These were taken at 4 approximate time points: 5, 12, 20 and 25 minutes after the start of each simulation. Participants verbally rated on a scale of 1 ("feeling no pressure") to 10 ("unable to cope with the pressure") how they were coping, when this information was requested by the Quality Assurance Officer.[10]

A final training room (H×L×W: 2.5m×6m×4m) was used to debrief participants and to complete a semi-structured interview after both simulations, which was cued by the presentation of a video of the participant completing the simulation. During the interview participants were asked to recall their thoughts about their decision-making at various points during the scenarios, their stress levels, and the application (or not) of operational discretion (see Appendices 6 and 7 for the questions). The responses to questions in the interviews were used *inter alia* to confirm the observed use of SOPs and operational discretion from the recordings of the simulations.

4.3.2.2 Simulation Scenarios. The two scenarios were designed and developed by two researchers who were recently retired, experienced incident commanders (advanced level commanders). The Discretion scenario was designed to replicate circumstances that required the application of operational discretion and use of professional judgement to make decisions. This simulation involved five young children who had fallen down a deep sinkhole in a remote location and included cues that related to each of the outcomes that justified (according to National Operational Guidance) the application of operational discretion: saving human life, taking decisive action to prevent an incident escalating, and where inaction may lead others to put themselves in danger. The standard operating procedure in this case is to enlist the support of specialist line rescue tactical advisers and teams to risk assess the situation and determine a plan to locate, rescue and recover them to

[10]Participants also wore heart-rate monitors, but these proved to be unreliable in approximately one third of participants.

the surface. This would involve securing additional specialist equipment and techniques to safely lower FRS and medical personnel into the sinkhole to assess the situation and condition of the casualties and to carry out their work to treat and recover them to the surface. However, embedded within the Discretion scenario were components that should have resulted in the use of operational discretion. The Control scenario involved a fire in a high-rise block of residential flats and included cues that informed the incident commander that the risks to life and property were low. As a result, the scenario was able to be successfully resolved with minimal risk to firefighters and the public by using the familiar SOP based on the service's generic risk assessment for firefighting in high rise buildings (Chief Fire and Rescue Service Adviser, 2014) and national operational guidance (N.O.G., 2021k): following a risk assessment, establish a bridgehead from which to launch a 2-line attack. That is, two breathing apparatus (BA) crews with hose lines, one to fight the fire in the flat, the other to protect their escape route from the lobby and be able to rescue them if necessary. This would take a minimum of six personnel (a Bridgehead Officer, BA Entry Control Officer and four BA wearers). There was no basis upon which to move beyond this SOP.

4.3.3 Procedure

Participants were tested between August and November 2019 and received one scenario in the morning and the second in the afternoon. In between the two scenarios, participants received lunch. Approximately half of the participants (21) received the Discretion scenario in the morning and the Control scenario in the afternoon, and the remainder (22) received the reverse arrangement. Immediately before and after each scenario, a blood sample was taken from one of the participant's fingers (and the LCC was assessed). Participants were then taken to the room in which the simulations were delivered. Before entering the room, the Quality Assurance Officer role-player gave the participants a general briefing on the time of year, day and the climatic conditions. They also read out a mobilising message from the mobilising control centre that outlined basic information about the incident. The participants were given an opportunity to ask questions of the mobilising control centre, as would be the case at real incident.

During both scenarios the incident commanders responded to the unfolding incident, complete with scheduled injects, in the way that they would a real incident: requesting information about resources and hazards, formulating plans, and directing the actions of their crews. On four occasions in each scenario they were asked to self-report their perceived level of stress by the Quality Assurance Officer. The responses of incident commanders to each scenario were video recorded and later scored as either using the requisite SOP (see Table 11) or departing from it and using operational discretion. Each video from all incident commanders was scored on two separate occasions to ensure the accuracy of the coding, and a subset of the videos were also scored by an independent researcher to confirm the reliability of the categorical coding. The semi-structured interviews took place after the final blood test following each scenario (see Appendices 6 and 7) and provided another basis upon which to confirm that operational discretion or an SOP had been applied.

Table 11: *Standard Operating Procedures and Examples of Operational Discretion*

Scenario	**Operational responses**
Discretion: Sinkhole rescue	**Standard operating procedure:** Enlist the support of specialist line rescue tactical advisers and teams to risk assess the situation and determine a plan to locate, rescue and recover the children to the surface. This would involve

	securing additional specialist equipment and techniques to safely lower FRS and medical personnel into the sinkhole to assess the situation and condition of the casualties and to carry out their work. **Operational discretion:** To save life, the committal of a firefighter down into the sinkhole using equipment designed to lower, but not raise, before the arrival of FRS specialist teams and equipment. **Operational discretion:** To save life, the committal of a Breathing Apparatus (BA) crew down into the sinkhole on two fully extended 10.5m ladders tied together before the arrival of specialist FRS teams and equipment.
Control: **High rise fire**	**Standard operating procedure:** Following a risk assessment, establish a bridgehead, two floors below the fire floor, from which to launch a 2-line attack. That is, 2 BA crews with hose lines, one to fight the fire in the flat, the other to protect their escape route from the lobby and be able to rescue them if necessary. This would take a minimum of six personnel (a Bridgehead Officer, BA Entry Control Officer and four BA wearers). **Operational discretion:** To prevent the situation from escalating the committal of a single line attack (i.e., a single BA crew with a hose line) from an established bridgehead, but without the required second BA crew to protect them.

4.4 Results

The overall results from the study were clear and are depicted in Figure 9A. Incident commanders were *less likely* to depart from using the SOP in the Discretion scenario, where such departures were licensed by the conditions, than in the Control scenario, where such departures were not licensed; examples of the use of operational discretion in the two scenarios can be found in Table 11. Thus, only five used operational discretion exclusively in the Discretion scenario, and a significantly greater number (18) used discretion exclusively in the Control scenario (binomial test, p =. 01); five used operational discretion in both scenarios and 15 did not use it in either (binomial test; $p < .05$). That is, there were more participants that did not use operational discretion at all than used it in both scenarios. McNemar's test confirmed that the proportions of the four types of response (represented by the four segments) differed (χ2 = 6.26, *p* = .01, OR = .28). Finally, the durations of the Discretion scenario (mean = 32.25 min; SEM = .45) did not differ significantly from those of the Control scenario (mean = 31.53 min, SEM = .66; *t*(42) = 1.05, *p* = .30, d = .158).

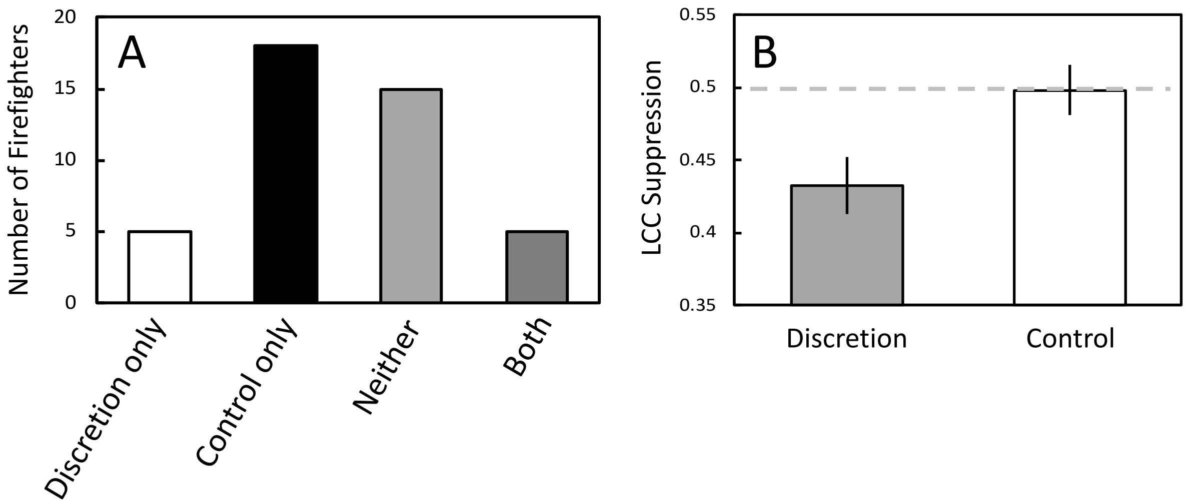


Figure 9 *Comparison Between The Frequencies of How Operational Discretion Was Applied by Incident Commanders (Panel A) and the LCC Suppression Scores (± SEM) Between the Two Scenarios (Panel B)*

Note: Panel A shows the numbers of incident commanders who exercised operational discretion in the two scenarios (Discretion and Control). Incident commanders were classified as using operational discretion in only the discretion scenario; in only the control scenario; in neither scenario; or in both scenarios. Panel B shows the mean suppression in LCC score, relative to baseline, after participation in the two scenarios.

The Discretion scenario resulted in more acute stress than the Control scenario, using both the blood-based assessment of leukocyte function (see Figure 9B) and self-reported stress. Figure 9B shows the mean suppression in Leukocyte Coping Capacity (LCC; ± SEM) after both scenario: LCC score after the scenario/(LCC score after scenario + before scenario).

As already noted, scores below .50 indicate a suppression in the LCC score after the scenario; with the degree of suppression indicating the capacity to cope with further stressors. The broken grey line indicates no suppression in the LCC score as a consequence of participation in the scenario. There was more suppression in LCC scores after the Discretion scenario than after the Control scenario ($t(42) = 2.206$, $p < .05$, d = .337); and one-sample *t*-tests confirmed that the scores for the Discretion scenario were below .50 ($t(42) = -3.391$, $p < .005$, d = .51), whereas those for the Control scenario were not ($t(42) = -.125$, $p = .902$, d = .02).

Table 12: *Mean (+SEM) Self-reported Stress at Four Successive Timepoints During the Two Scenarios*

Timepoint	1	2	3	4
Discretion Scenario	3.81 (0.24)	5.30 (0.24)	5.55 (0.22)	6.58 (0.20)
Control Scenario	3.86 (0.26)	4.79 (0.25)	5.44 (0.25)	6.16 (0.22)

The self-reported stress scores (minimum = 0 and maximum =10) increased across both scenarios (see Table 12). ANOVA revealed no main effect of scenario, $F(1, 42) = 1.473$, $p = .232$, $np^2 = .034$, a main effect of test (1-4), $F(3, 126) = 101.686$, $p < .001$, $np^2 = .708$, and an interaction between these factors, $F(3, 126) = 2.671$, $p = .05$, $np^2 = .060$; with the scores for the Discretion scenario being higher than the Control scenario on test 2 ($t(42) = 2.14$, $p < .05$, d = .32). There was a *negative correlation* between the final self-reported stress score (high scores = more stress) from the Discretion scenario and the raw LCC scores (low scores = less residual capacity to cope with stress) taken after the scenario, using both Pearson's (r_p) and Spearman's (r_s) correlations ($r_p = –.446$, $p < .005$; $r_s = –.439$, $p < .005$); but there was no correlation for the corresponding scores for the Control scenario ($r_p = -.036$, $p = .82$; $r_s = .051$, $p = .74$). Given the fact that the Control scenario did not result in a suppression of LCC scores the latter observation is not particularly surprising.

The chronic stress scores taken from the questionnaire (mean = 6.18, SEM = 0.29) did not correlate with the measures of acute stress during either scenario: LCC ratios from Discretion scenario (r_p = -.08, p = .57; see Figure 9B for means), Control scenario (r_p = .06, p = .72; see Figure 9B for means); or with self-reported stress scores on test 2 in which the Discretion and Control scenarios differed (r_p = .18, p = .25 and r_p = .25, p = .11, respectively).

Of the 43 incident commanders, 32 were intermediate level and 11 were advanced level. The proportions of participants at the two levels who only used operational discretion in either the Discretion scenario or the Control scenario was consistent with the overall pattern of results depicted in Figure 1A: Intermediate level (4 versus 14) and Advanced level (1 versus 4). These proportions did not differ using a Fisher's exact probability test ($p > .05$). However, there was some indication that the proportions that used operational discretion in neither or both scenarios differed between the levels: Intermediate level (neither = 13 versus both=1) and Advanced level (neither = 2 versus both=4; $p < .05$). This difference, albeit with a very small number of advanced level incident commanders, suggests that a general reluctance to use operational discretion was more evident in the intermediate level commanders than the advanced level commanders. A secondary analysis of the LCC suppression scores for the two scenarios, including the two command levels, revealed a similar pattern to that depicted in Figure 9B: intermediate = .44 (Discretion; SEM =.023) and .49 (Control; SEM=.016); and advanced = .40 (Discretion; SEM=.037) and .51 (Control; SEM=.051). ANOVA revealed no effect of command level, $F(1, 41) = .181$, $p > .67$, $np^2 = .004$, an effect of simulation, $F(1, 41) = 5.59$, $p = .02$, $np^2 = .12$, and no interaction between these factors, $F(1, 41) = 0.83$, p > .36, np^2=.02.

4.5 Discussion

Decisions made by firefighters can mitigate the economic, environmental and social impacts of emergency incidents. Guidance given to firefighters in the UK Fire and Rescue Service recognises two approaches to decision making with the recommendation that responses to routine emergency incidents are based on rules (i.e., standard operating procedures, SOPs) and 'unprecedented' incidents licensing the use of a more flexible, deliberative approach (i.e., operational discretion). The recognition of these two processes is echoed in psychological theory, where the use of rules (Epstein, 1994; Gigerenzer, & Goldstein, 1996; Klein, 1993; Sloman, 1996) is distinguished from a deliberative approach involving a cost-benefit analysis (Kahneman, 2003; Von Neumann & Morgenstern, 1947). My research concerned the deployment of SOPs and use of operational discretion in experienced firefighters. I used two scenarios: The Discretion scenario licensed the use of operational discretion (e.g., on the basis of saving human life) and the Control scenario did not. The use of operational discretion was neither random (equally evident in both scenarios) nor was it consistently used by different incident commanders (either always using it or never doing so). In fact, only 5 of the 43 firefighters used operational discretion in a scenario-appropriate manner.

If we first consider the Control scenario alone: a fire in a flat in which there is no danger to human life. This is a relatively routine incident, for which the SOP is well established (see Table 11) and participants were informed that there was no risk to human life; and yet over half of the participants (23) used operational discretion without justification for doing so. This observation is, in and of itself, is important. The Discretion scenario was less routine, and there was a clear risk to human life involving the children who had fallen into a sinkhole. Taking this scenario alone, only 10 of the 43 participants used operational discretion. The difference between the scenarios may explain the significant difference in the level of self-reported stress on the second occasion when the question was asked. The timings of the stress self-report questions were the same for both scenarios but the development of the incidents varied between incident commanders depending on their decision making. However, given the timing of the second stress self-report question (12 minutes) it is perhaps not surprising to find a significant difference as it occurs a short time after the incident commanders have gathered their initial information and determined the status of the incident. In the case of the Discretion scenario, it would have been recently confirmed to the incident commander that five young children have fallen 20m into a sinkhole. In contrast, in the Control scenario they would have recently found out that no lives are at risk and the fire is contained.

If the two scenarios are now taken together, there were more incident commanders who used operational discretion in the Control scenario and not in the Discretion scenario than incident commanders who used operational discretion in a context-appropriate manner. Whether the results of the scenarios are taken separately or together the results have important implications for firefighter training. These observations themselves prompt two questions: Why was operational discretion used when an entirely appropriate SOP was available? Why when the conditions licenced operational discretion was it not used? One possibility is informed by the fact that the Discretion scenario generated greater acute stress than the Control scenario, as measured by both immune function and self-report.

The results of laboratory studies show that extrinsic stress can result in a reliance on rules rather than deliberation (Kassam et al., 2009; Starke & Brand, 2012; see also, Janis &

Mann, 1977; Peters, McEwen, & Friston, 2017). By the same token, the fact that the Discretion scenario generated more stress than the Control scenario might then have generated a greater reliance on SOPs than operational discretion. But how is acute stress generated and how does it impact decision making? One influential class of psychobiological accounts assumes that acute stress is generated when the perceived demands of the situation are judged to be beyond the personal and environmental resources that are available to address those demands (see Lazarus & Folkman, 1984; Salas, Driskell & Hughes, 1996; for a review, see Mark & Smith, 2008). To the extent that the Discretion scenario involves such a mismatch, including the grounds for the use of operational discretion, then it would be expected to generate acute stress. There are a variety of plausible mechanisms by which acute stress – generated in this way – could affect the use of SOPs and operational discretion. For example, it could limit attentional resources and thereby constrain either (i) the capacity for the deliberative processes upon which operational discretion relies (e.g., Combs & Taylor, 1952; Easterbrook, 1959), or (ii) the requisite situational awareness (Endsley, 1995).

The results of Study 3a were clear. However, while assessing whether or not operational discretion was used is important, it does not help to understand how the use of command skills varies between the two scenarios. Moreover, it does not help us to understand how those incident commanders who used operational discretion in a context appropriate manner differed from those who did not. Study 3b addressed these two issues through a THINCS assessment of command skills exhibited by the incident commanders during the two scenarios, and a qualitative analysis of the content of the semi-structured interviews with them.

Chapter 5

Study 3b: An Investigation into the Different Uses of Command Skills and Operational Discretion by Incident Commanders

5.1 Abstract

Chapter 5 presents further analyses of the use of operational discretion by of a subset of participants from Study 3a described in Chapter 4. Briefly, the participants responded to two scenarios, one that licenced the use of operational discretion (the discretion scenario) and one that did not (the control scenario). The first component of the analysis used THINCS to assess the six command skills during the two scenarios, and the second component involved a (deductive) thematic analysis conducted on the subset of the participants who used operational discretion in a context appropriate manner, and a control subset who did not. The primary aim of these additional analyses was to gain a more complete understanding of the basis of their use of operational discretion.

5.2 Introduction

Chapter 4 described Study 3a in which incident commanders responded to two scenarios, one of which licensed the use of operational discretion and another that required the use of an SOP. Their responses to both scenarios were categorised as involving the use of operational discretion or not. There were only 5 participants who used operational discretion in a context appropriate way, with significantly more participants (18) using it in a context inappropriate way. Study 3a also revealed that the scenario that licenced the use of operational discretion generated more acute stress as measured both using a blood-based assessment and self-report. These results are important and suggest that stress might well result in paradoxical use of operational discretion (and SOPs). This study aimed to gain a more detailed understanding of the command skills deployed in the two scenarios and the origins of differences between the appropriate (and inappropriate) use of operational discretion and SOPs. The first part of the analysis used THINCS to assess the six command skills during the two scenarios in a subset of 20 incident commanders. The second part of the analysis was a (deductive) thematic analysis conducted on the subset of the participants (*n*=5) who used operational discretion in a context appropriate manner, and a control subset who did not (*n*=5). This analysis was conducted on a sub-set of the questions that followed each scenario. Braun and Clarke's (2006) method of thematic analysis was chosen to analyse the transcripts of the participants responses to these questions. Alternative methodologies such as conversation analysis (see Hutchby & Wooffitt, 1998), interpretative phenomenological analysis (see Smith, Flowers & Larkin, 2009) or discourse analysis (see Potter & Wetherell, 1987) are tied to specific theoretical or epistemological positions which limits their flexibility (Braun & Clarke, 2006). Compared to these, Braun and Clarke's (2006) thematic analysis is theoretically flexible because they advocate it is not necessary to follow a theory or explanatory framework to find and analyse patterns in language (Braun & Clarke, 2013).

5.3 Method

5.3.1 Participants

The first part of this study involved 10 advanced level commanders and 10 intermediate level incident commanders from Study 3a. The second part involved the five incident commanders who used operational discretion in a context appropriate manner and a random sample of five who used operational discretion inappropriately.

5.3.2 Materials

For the initial part of Study 3b the THINCS app (see appendix 2) was used to rate the command skills of the 10 advanced and 10 intermediate level incident commanders based upon video footage of their performance in the two scenarios: Discretion and Control. The second part of Study 3b used audio recordings of the interviews following both scenarios to produce transcripts for the thematic analysis of the five incident commanders who applied operational discretion appropriately and the five who did not.

5.3.3 Procedure

The THINCS assessments, interview transcription and thematic analysis were solely conducted by PCB. Due to PCB's level of subject matter expertise in THINCS rating, use of the app and the lack of other subject matter experts meant that a second Rater was not employed to verify PCB's THINCS ratings. For similar reasons a second researcher was not used to cross-code the outcomes of the thematic analyses.

The THINCS assessments were conducted using the THINCS app. PCB paused the incident commanders' video footage to optimise the recorded number of behavioural observations and ensure accurate ratings were awarded. The transcripts of the interviews focused on the answers to the following questions (see Appendices 6 & 7) from the interview:

- Initial understanding of incident
- Identified hazards
- Initial plan including objectives and priorities

- Why a BA team was/was not committed from the bridgehead (Control scenario only)
- Use of the Aerial Ladder Platform (ALP; Control scenario only)
- Use of Urban Search and Rescue (USAR) line rescue adviser (Discretion scenario only)
- Use of medical Hazardous Area Response Team (HART; Discretion scenario only)
- Self-reported stress
- Application or not of operational discretion

The transcripts were then subject to a thematic analysis by PCB using an adaptation of the procedure described in Braun and Clarke (2006) using NVivo software (QSR International, 2018). Braun and Clarke (2006) state that thematic analysis may be done deductively or inductively. For this study a deductive thematic analysis was completed as I was interested in specific details contained within the interview transcripts, not just what was 'grounded' in the data. The lens through which PCB carried out their deductive thematic analysis was from the perspective of a former firefighter with 31 years of experience over ten years of which was as an advanced level incident commander. Also, as someone who studied qualitative research methods at undergraduate and postgraduate degree levels. Finally, as someone who has carried out decision making research using both quantitative and qualitative methods.

Braun and Clarke (2006) describe a 6-stage step by step guide. The 6 stages are:

1. Data familiarisation (reading and transcribing data)
2. Initial coding (systematically coding the data and collating data for each code)

3. Theme identification (Collating the codes into an initial set of themes)

4. Theme review (Reviewing and finalising the set of themes)

5. Theme definition (Further analysis to set and define each theme)

6. Report writing (Selection of illustrative examples for each theme and compilation of an academic report)

The adapted version conducted by PCB did not fully complete the final two stages. The themes were not defined as the deductive process aligned them with the predefined command skills and the findings have not been the subject of a separate academic report.

5.4 Results

5.4.1 The use of (THINCS) command skills during the two scenarios

From the 43 incident commanders a sample of 20 were randomly selected and placed into two equal groups that were counter-balanced by their level of command and as much as possible by their use of operational discretion. Figure 10 shows the mean scores for each of the six THINCS skills in the two scenarios. Inspection of this figure reveals that there were differences between the scores for the two scenarios in effective decision making and planning, and in situational awareness. An ANOVA was conducted with scenario type, level of command, and command skills as factors was conducted confirmed this description of the results. There was no main effect of command level ($F(1, 18) = 1.74$, $p = .204$, $np^2 = .008$) or scenario type ($F(1, 18) = 3.241$, $p = .089$, $np^2 = .088$), but there was a main effect of command skill ($F(5, 90) = 8.632$, $p < .001$, $np^2 = .324$) and an interaction between scenario and command skill ($F(5, 90) = 4.969$, $p < .001$, $np^2 = .216$). There were no other significant two-way interactions or a three-way interaction (largest $F(5, 90) = 1.18$, $p = .324$, $np^2 = .062$).

To explore further the interaction between scenario and command skill a series of *t*-tests were conducted. The difference in effective decision making and planning was not statistically significant ($t(19) = 2.059$, $p = .055$, $d = .458$), but there was a difference in situational awareness ($t(19) = 5.667$, $p < .001$, $d = 1.267$). There were no other differences.

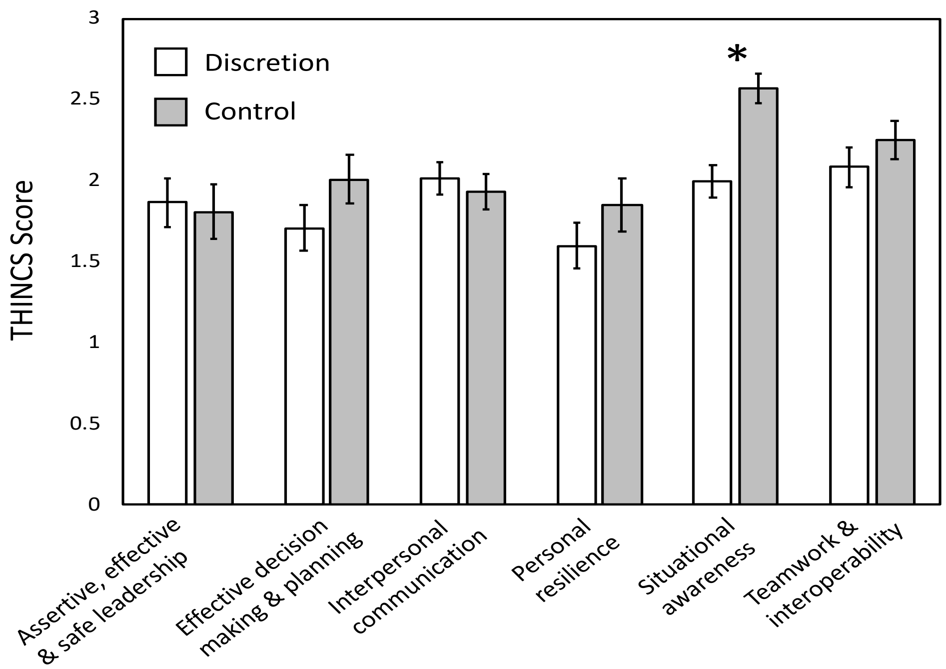


Figure 10: *Mean (±SEM) Scores for the THINCS Commands Skills During the Discretion and Control Scenarios (* = p < .001).*

5.4.2 Thematic analysis of the appropriate and inappropriate use of operational discretion

A total of 20 interviews were transcribed: The appropriate (A) group consisted of the five commanders who applied operational discretion appropriately across the two scenarios, whereas the inappropriate (I) group consisted of five randomly selected participants who applied operational discretion during the Control scenario and did not apply operational discretion in the Discretion scenario. The thematic analysis focussed on the responses to key questions during the semi-structured cued recall interviews recorded after the two scenarios (see Section 5.3.2 above). These answers were subject to thematic analysis (adapted from Braun & Clarke, 2006) and using NVivo software (QSR International,

2018) to produce a set of themes for each group for each scenario (see Table 13). These themes were then refined and categorised using the THINCS Command Skills (Tables 14 and 15).

Table 13: *Identified Themes From the Cued Recall Interview Transcripts of Incident Commanders Who Used Operational Discretion Appropriately (Group - A) and Those Who Did So Inappropriately (Group I).*

Control Scenario		Discretion Scenario	
Group A	**Group I**	**Group A**	**Group I**
Competence	Competence	Competence	Confidence
Culture	Culture	Confidence	Culture
Operational discretion	Decision making	Culture	Operational discretion
Planning	Operational discretion	Operational discretion	People oriented
Priorities	People oriented	People oriented	Planning
Risk	Planning	Planning	Priorities
Safety leadership	Priorities	Priorities	Risk
Setting standards	Risk	Risk	Safety leadership
Situational awareness	Safety leadership	Safety leadership	Setting standards
Stress	Setting standards	Setting standards	Situational awareness
Teamwork	Situational awareness	Situational awareness	Stress
Uncertainty	Stress	Stress	Teamwork
Values and supports others	Teamwork	Teamwork	Uncertainty
	Uncertainty	Uncertainty	Values and supports others

The themes identified from the interview transcripts for group appropriate (A) and inappropriate (I) appear to be relatively similar. However, examination of the illustrative quotes in Tables 13 and 14 highlight underlying differences between them. Only themes that included more than 3 or more references from each group were compared.

5.4.2.1 Summary of the control scenario (Table 14). There were similarities with respect to leadership (competence), personal resilience (stress and uncertainty), and teamwork and interoperability (teamwork). However, there were also key differences between the groups in relation to leadership (priorities, risk, safety leadership, and setting standards), decision making (operational discretion and planning), and situational awareness. Incident commanders from both groups demonstrated their competence with respect to the high rise SOP, such as the minimum requirement for a 2-line BA attack consisting of two BA crews of two and two lines of firefighting hose. Also, the incident commanders reported similar stressors (e.g., that caused by limited and delayed resources). With respect to uncertainty, they shared similar concerns about the potential involvement of people in the fire. However, there were differences between the groups: The situation awareness of incident commanders who inappropriately applied operational discretion (those in Group I) was dominated by an errant anticipation that the fire would spread. They also, planned and prioritised an offensive attack on the fire above firefighter safety and judged the risks to be sufficiently low to do so without the requisite resources in place. In doing so, they demonstrated inadequate safety leadership and set a poor standard of performance by making a decision to use a single-line BA attack on the fire (operational discretion) based on repeated requests from the bridgehead officer. In contrast, incident commanders who applied the requisite SOP (those in Group A) understood the limited potential for fire spread and showed a different approach to Group I. They prioritised firefighter safety by planning a defensive attack on the fire from the outside of the building using an Aerial Ladder Platform (ALP – a high rise firefighting vehicle) having judged the risks to be too high to permit an internal attack on the fire with insufficient resources. They consistently refused to allow a single-line BA attack despite several requests to do so, showing effective safety leadership and setting the correct standard of performance by following the SOP.

Table 14: *Comparison of Groups of Incident Commanders Who Showed Appropriate (Group A) or Inappropriate (Group I) Use of Operational Discretion in the Control Scenario; with Key Themes Illustrated by Quotations From the Transcripts.*

Command Skill Themes	Group	Refs	Illustrative Quotations
Assertive, effective and safe leadership			
Competence	A	12	I...ask...for the [high rise] SOP and everything to be printed off. I've got it in my head, you know, roughly in my head anyway the SOP for the 2,3-line attack and everything else.
	I	3	[With respect to the] high rise SOP…you've got 2-line or 3-line [attacks] then, so, there are certain circumstances in which the 2-line is specified. If you know that there is somebody in there and you've had it confirmed they're on the line to Control or something like that, you know. There's also in [the SOP] for 2-line for preventing significant escalations, so you could justify the 2-line under the procedure...not with a 1-line, a 1-line does not exist.
Priorities	A	13	[Priorities were] obviously save life and, but, you know, to, at that point, I was, you know, the information I'd had, there was no persons reported [missing] in the [fire] flat so, the second was obviously, you know, to attack the fire.
	I	8	My priority was to fire-fight, prevent escalation.
Risk	A	23	If we'd have gone [into the fire flat] and put the fire out quite quickly then the evacuation wouldn't have needed to have continued. We could have dealt with the incident quite quickly. I was concerned about escalation, you know, once you get committed into a building like that without any support, not even a safety jet in place, that it was too higher risk to take, bearing in mind there was no report of persons in that part of the building.
	I	18	[With] a significant delay on the oncoming resources, it made more sense to accept an amount of risk in order to try and put some water on the fire and damp it down. At least, maybe you couldn't extinguish it, but if you could knock it down and buy yourself some time for us then to be able to put 2 or 3 lines in, you know, as soon as the resources allow.

Safety leadership	A	34	[Decision not to commit a 1-line attack because] we didn't have a safety jet, a means of providing [the BA crew] with any support or protection. They were on their own and I think that concern about skill and familiarisation was in my mind there. No life risk. Do I commit for property? I was, I think happy with the fire in that compartment contained.
	I	19	The information coming back from the [bridgehead and sector commanders] saying we can deal with it, we've told you not once, its not a snap decision, we've told you 3 times now...and that's like OK we're going, we're doing it
Setting standards	A	7	Still [following] the high rise SOP, that's all we've, regarding that. It's what I'm following.
	I	11	I think probably [I felt I was operating] within [the high rise SOP], I am working towards having full policy in place, of grasping that window of opp, erm, sorry to keep using the phrase, keep grasping that window of opportunity [to attack the fire].
Effective decision making and planning			
Operational discretion	A	11	No, [not comfortable with ops discretion] that would be a very fully considered decision and I think not one that you would make lightly, one that you would have to provide a good rationale for as to why you did it. And I think you probably need some more information directly from [the relative in the flat above the fire]. If she was in the flat directly above the fire and was being effected by smoke, then I think you may want to utilise what resources we had with what [HART] had and [rescue them]......So, you're still not committing [a BA crew] to the fire compartment, but sending the firefighters you've got up there with whatever capability they've got to provide as much support as they could [to HART], almost like a snatch rescue scenario, you know, go in, get out, come back out.
	I	25	Delays in oncoming resources [influenced my decision to use ops discretion and] the evident impression that was coming back from the crew that they were confident to make an entry. So, there're in the best position to, you know, make an assessment of the conditions and decide if they think that its suitable.
Planning	A	14	My thinking now is to get the [bridgehead] crews to not attack the fire on the 11th but to proceed to the 12th and get the [FSG residents] out, that's what I was thinking, you know, its not attacking the fire but leave everything in place to go to the 12th evacuate those flats because that then gets them out the way while the ALP is still attacking the fire.
	I	13	I'm thinking, in light of the short number of resources I've got I want them focused on securing the attack rather than going above the fire and potentially putting themselves at risk saving these people.

Personal resilience			
Stress	A	26	I think its from past, like what we was saying [knowledge of others'] experiences of these types of fires. We know they're very difficult to fight. We know that people, crews will do things that they shouldn't do because they think they're doing the right thing, put themselves in danger. And I had very limited resources and [oncoming] resources were going to take a very long time to [arrive]. So, in my mind I'm thinking I've only got 8 people to try and put all this out and if people start self-evacuating and all sorts of stuff I'm in a world of damage here.
	I	11	It's just because you're on your own. You know its getting worse. All the time we're doing nothing, its getting worse and worse. And it’s just not, it’s never going to happen as quick as you want it to. So yeah, for me, it was starting to get a little bit, OK, this is getting, the longer it goes on the more involved we're going to get in it. So, that’s kind of, its always at the back of my mind is, the longer we're here doing this without any ALP we're exposed and so are the people in [the block of flats].
Uncertainty	A	21	I'm thinking [the oncoming resources] are going to be a little while, what if there is someone in there, what if it does escalate? We're going to have to do something else.
	I	20	We're not 100% sure if its persons reported or not. We've got information there is no one in there, but we've had no one confirm that with us, so that for me is the big one. [What] is the state of the fire, what [IC1 has] briefed his teams to do, and is there any one in [the flat]?
Situational awareness			
Situational awareness	A	47	I feel with limited resources you have to, if you've got 2 pumps, you've got 2 pumps. You can't magic pumps up…It was a flat fire, yeah I could see it was a flat fire, it was a going job, [I could] see it was, it was where it were, [IC1] said it was, 11th floor. I was happy it wasn't a cladding building, so I wasn't too fussed about rapid fire, it was behaving as I would expect it to behave. I wasn't, I know resources were coming. I was pretty chilled out about it.
	I	29	So, delayed resources...what I can see from the outside is that we don't have flames coming out of the window, I can see that we haven't got a wind drive situation. So, the risks there are relatively low, the benefit is now potentially quite significant because we know we've got fire survival guidance advice being given to people above the fire, so really, for the stay put [policy] to work, you've got to extinguish the fire.
Teamwork & interoperability			
People oriented	A	0	

	I	2	It wasn't that I didn't trust IC1 it was more that I felt that he was perhaps...he was either a little bit overwhelmed maybe or not fully au fait with what he was going to set in there.
Teamwork	A	19	[I made IC1] Fire SC...because that's where the information was. So, its pointless me putting him in the Lobby, it pointless anywhere else other than that's where the information is. So I had, and I couldn't tell him to go to the 11th floor, so 9th floor, you assess it, you, when you're up there. So, so, it's based a lot on trust is that, you know [that] they're going to give you the information.
	I	10	Let's get his expertise, he's a WM, let's get him up there as the fire sector commander. I want someone up there who's going to make a difference early doors 'cause I'm limited.

Note Text in [brackets] has been inserted to clarify the context of a quotation and typos have been corrected

5.4.2.2 Summary of the discretion scenario (Table 15). There were similarities between the groups with respect to teamwork and interoperability (teamwork) and personal resilience (uncertainty). However, there were key differences in relation to leadership (priorities, risk, and safety leadership), decision making (operational discretion and planning), personal resilience (confidence and stress), and situational awareness. Incident commanders from both groups formed teams that put together the line rescue expertise of USAR tactical advisers and HART team leaders, and they expressed similar uncertainties with respect to the aspects of the scenario such as the wellbeing of the children. However, situational awareness in incident commanders in Group I reflected an acceptance that they were unable to act until all the specialist personnel and equipment were in place. They also prioritised firefighter safety above the lives of the children, planned a defensive approach until a rescue plan was decided upon by the line rescue specialists, and judged the risks to be sufficiently high enough to prevent a rapid intervention with the available resources. In doing so, they demonstrated inadequate safety leadership by deciding not to use operational discretion to access the sinkhole despite it being licensed by the circumstances. Incident commanders in Group I demonstrated less confidence in relation to their decision making about the application of operational discretion, and their stress was influenced by limited line rescue expertise and the lack of a rescue plan. In contrast, situational awareness in incident commanders who applied operational discretion (those in Group A) suggested they were more focused on the welfare of the children. They showed a different approach to Group I as they recognised they could take action with the immediately available resources before the arrival of the specialist equipment and personnel. They prioritised the safety of the children over that of the firefighters and planned a rapid, offensive intervention to rescue the children as they judged the risks to the children to be higher than those to their firefighters. They demonstrated effective safety leadership by being prepared to accept the additional risks to which they would be exposing their firefighters in order to save human lives, which was also a main source of stress.

Table 15: *Comparison of Groups of Incident Commanders Who Showed Appropriate (A) or Inappropriate (I) Use of Operational Discretion in the Discretion Scenario; With Key Themes Illustrated by Quotations From the Transcripts.*

Command Skill Themes	Group	Refs	Illustrative Quotations
Assertive, effective and safe leadership			
Competence	A	5	[This incident] was more a mixture between trench [rescue] and working at height.
	I	0	
Priorities	A	13	To get safely someone down there to assess those casualties.

	I	10	The priorities, were scene [and firefighter] safety and information gathering from the children whilst awaiting the arrival of line rescue experts and then formulate a plan to lower people into the hole and then figure out a plan about how we are going to get everybody out.
Risk	A	13	You are going to be limited with the control measures you put fully in place for [a firefighter], you'd be heavily relying on brute strength to [recover them from the sinkhole], but the need for it, the severity of the casualties, once you know they are unconscious you've got no options.
	I	17	Withdraw as many firefighters as we can from the immediate vicinity of the hole, so that was slightly of concern to me because they weren't doing anything, and if you're not doing anything what's the point of being in hazard zone? Not only are they putting themselves at risk they are potentially putting the kids in the hole at risk as well because if we do get any further collapse, you know, a clod of earth that falls 20m onto your face is gonna hurt.
Safety leadership	A	11	We were using other agencies [HART] to create, to improve the safe systems of work, but it doesn't stop us from...from the initial action of getting somebody into the hole just [based] on the...you know, on the condition of the children.
	I	8	Let’s not, let’s not be cavalier, let’s not, you know, sort of put people unduly at risk, but actually I think we need to do something and that probably needs the visual leadership of me there to kind of do that
Setting standards	A	2	I should have really [declared ops discretion] because operating outside of operational procedures should be notified to [mobilising] Control.
	I	8	[Didn't implement ops discretion because] we were in the process of getting it sorted and never really got the opportunity because we had the full line rescue capability arrive and we were working within [SOPs].
Effective decision making and planning			
Operational discretion	A	18	When [it was reported that] we had lost contact with all of [the children], then that might have triggered something, right we've got to get someone down there, I want to know what's going on because we've got no contact at all. It might be, you know, we've got to get them out now, sort of thing. So yeah, it was all about ops discretion.
	I	17	The devil on my shoulder was doing an out of procedure, out of protocol type of rescue or type of procedure, and to be pushed into using the op.....I think the phrase I'm using there is being pushed into making operational disc...using operational discretion is because I'm forced either because of the dynamics or the timescale of the incident.
Planning	A	15	The plan was to get someone down [in the sinkhole] then as they were down there to get everything else set so we could actually start lifting [the children] out and to keep the old parents and everything away.

	I	20	I'm the incident commander I own the plan, I need to know what it is. I'm happy for them [USAR tactical adviser and HART team leader] to make a decision on what they're gonna do, I want to know what it is.
Personal resilience			
Confidence	A	4	At the end of the day I know its a risk to the firefighters, but we're there to rescue people and its probably firefighters want to do it anyway and they need to know that I've, I've, they've got my backing...that's my decision.
	I	3	When the risk massively goes up is then you know I'm not necessarily going to take that decision in isolation. I might be taking it in isolation in terms of I'm the commander there at that moment in time, but actually I'd be consulting.
Stress	A	8	[Stress was born out of] knowing that I'm...stretching my...like I said, you know, going out of procedure to enact the rescue, but, but, but also knowing that, you know, what I, what I'm conscious about is the risk to my people.
	I	14	I'm putting my faith in people. I need to know what are they doing? And its rightly so, because they've got the expertise I don't have so I need them, but I need to know what are they doing? I own the tactics, I'm the incident commander, I own the risk, I need to make sure I'm happy.
Uncertainty	A	13	I wanted to…find out how deep [the sinkhole] was and the level of injuries, you know, what, if [the children are] happy down there? Then its slow, slow, slow. If they're dying? Quick, quick, quick.
	I	11	We didn't know about any services or utilities; we didn't know about the air quality in the hole. We didn't know what injuries the kids had sustained...We didn't know once USAR [line rescue team] had arrived how long it would take to set up a safe system to access [the sinkhole] and extricate [the children]...and we didn't know really how stable the ground was around the sinkhole.
Situational awareness			
Situational awareness	A	27	The hazards really again, are its raining now, its soft ground, so for us its caving in [of the sinkhole edges] which would have an effect on the kids by being buried...this, this has to be a quick rescue, this is not an hour or two rescue, this needs to be pretty quick.
	I	25	It’s not like it’s a cave where people can get themselves to a position of safety potentially where we've got the option to mount, you know a week-long rescue, we've got kids down there, they're unconscious, there's very little we can do to mitigate the risk other than to get the technical specialist teams in.
Teamwork & interoperability			

People oriented	A	2	[The initial incident commander] was giving the impression that he wasn't comfortable in that position as being exposed to it. He was running out of ideas and wasn't sure how to deal with the incident and basically, put in the control measures needed to get…information.
	I	1	The [initial] incident commander at the time seemed nervous, out of depth, all of those things so I needed to take that pressure of him 'cause I needed him to be operational.
Teamwork	A	6	[In my mind I think] that the Operations Commander and the [USAR] tactical adviser are going to work together to put a firefighter in the hole.
	I	13	[HART team leader has] told me what I want to know really. What her priority is, why she's there, we've got a common goal obviously getting down to the kids and they've got the capability to do stuff. So again, make sure she is aware of what we've got, get her down [to the sinkhole], get her talking with our [USAR] expert, let's get a plan. We've got stuff we need there to do, you know, meaningful stuff.

Note Text in [brackets] has been inserted to clarify the context of a quotation and typos have been corrected

5.5 Discussion

The aim of Study 3b was to provide a further analysis of the basis for the use of operational discretion observed in Study 3a. One component of this analysis relied on the use of THINCS to assess how the use of the command skills differed between scenarios where operational discretion was either licensed or not. The assessment was conducted on a sample of 20 incident commanders from Study 3a. The principal difference between the two scenarios was in the situational assessments that formed the basis of situational awareness (Sarter & Woods, 1991). The resulting situational awareness was lower in the discretion scenario than in the control scenario. This is an interesting observation, when coupled with the fact that there were no statistically significant differences in the other command skills. It provides converging evidence for one interpretation of the results reported in Study 3a: The suggestion that stress impacts situational awareness upon which the use of operational discretion relies (cf. see Section 4.5; cf. Endsley, 1995). Situational awareness involves the cognitive processes of perception, attention and working memory (Endsley, 1995) all of which are known to be adversely affected by stress (for a review see Staal, 2004). Couple this with the findings from studies on firefighters that showed they found incidents involving children to be the most stressful (Baker & Williams, 2001; Haslam & Mallon, 2003) and it is perhaps not surprising that the situational awareness of incident commanders between the two scenarios was different. However, this does not explain why incident commanders inappropriately used operational discretion in the Control scenario. An explanation for this may lie in the role of expertise and the access experts have to a broad variety of mental models (patterns) that they use to interpret situations (Endsley, 1995; Sarter & Woods, 1991). Future research should examine the impact of incident command experience in terms of its variety and depth upon situational awareness (e.g., where information is limited or missing).

The second component of Study 3b was a qualitative analysis of the differences between five incident commanders who used operational discretion in a context appropriate manner and five who used it in a contact inappropriate manner. This analysis provided insights into why operational discretion was used (cf. Study 3a). For example, the analysis of the responses from Group I suggested that the inappropriate use of operational discretion (in the control scenario) reflected the anticipated spread of the fire; and their failure to use operational discretion (in the discretion scenario) was associated with the stress of not having the necessary resources to rescue the children involved using the SOP. In contrast, the appropriate use of an SOP in Group A was associated with less concern about the potential for fire spread and consistently refusing to allow a single-line BA attack in favour of a defensive attack on the fire using the ALP; and their use of operational discretion in the discretion scenario reflected the safety of the children and a willingness to accept proportionate risks to firefighters. The appropriate deployment of operational discretion seems to reflect an evaluation of costs and benefits.

Situational awareness is comprised of the rapidly changing information born of situation assessments which dictate the access to knowledge and experience that forms the basis of decisions (Endsley, 2021). In the case of Group I, the decision to use operational discretion in the Control scenario might have been due to a lack of knowledge and/or experience in commanding high rise building fires that resulted in an assessment of the situation that anticipated fire spread. The stress of commanding an incident involving children (Baker & Williams, 2001; Haslam & Mallon, 2003) and the pressure to use professional judgement to make decisions beyond the scope of SOPs may account for the poor situation assessments that led to operational discretion not being used in the Discretion scenario. In the case of Group A, the decisions not to use operational discretion in the Control scenario and to use it in the Discretion scenario could have been as a result of situation assessments based on greater knowledge and/or experience with respect to commanding high rise building fires and incidents involving a variety of rescue techniques, e.g. from height, trenches and building collapses.

The results of Study 3b complement those of studies 2a, 2b and 3a, providing further evidence that THINCS is sensitive to differences in command skills (i.e., situational awareness), differences which were reinforced through a qualitative analysis of post-scenario interviews. In Chapter 6, I consider the broader implications of the results of studies 3a and 3b for the psychology of incident command.

Chapter 6:

General Discussion and Implications for Policy, Practice and Research

6.1 Summary of objectives

The UK Fire and Rescue Service is a high reliability industry in which human error can have catastrophic consequences. Other high reliability industries have adopted the approach of developing behavioural marker systems to evaluate the non-technical (i.e., psychological skills), which can be used in the context of training and evaluation. No such system exists for the UK FRS. This thesis had two objectives relating to this issue: To develop a behavioural marker system for incident command in the UK Fire and Rescue Service; and to evaluate the resulting behavioural marker system. Without such a system, it is also difficult to assess the impact of experimental manipulation on the presence or otherwise of the requisite command skills. The third objective of this thesis was to use the resulting behavioural marker system in a research context, to study the use of SOPs and operational discretion in different simulated emergency scenarios.

6.2 Summary of findings and impact

Study 1 describes the development of behavioural markers for the THINCS command skills (Butler, 2016; Butler et al., 2020) using workshops with representatives of the UK FRS (i.e., subject-matter experts). It consisted of six core command skills (Flin et al., 2008), with 20 subskills bespoke to the UK FRS context, each with their own behavioural indices and rating system. As part of this work, an app was also developed. Studies 1, 2a and 2b also involved an evaluation of the system's psychometric properties, including its reliability and validity, which proved to be acceptable. The app was developed with MyOxygen Ltd with additional ESRC Impact Acceleration Account funding and is now being deployed widely within the UK FRS and is featured in National Operational Guidance (N.O.G., 2021l). The evaluation of the system presented in Chapter 3 received additional funding from the Fire Service Research and Training Trust. The front end of the app is depicted in Figure 11 (see also Appendix 2).

Figure 11: *The Pump Impeller Graphic Displayed on the Front of the THINCS App*

Studies 3a and 3b were co-produced with the National Fire Chiefs Council and Hampshire Fire and Rescue Service to examine the conditions under which SOPs and operational discretion are used in two simulated scenarios. Study 3a showed that these two approaches to decision making were not deployed in the ways anticipated by National Operational Guidance (N.O.G., 2021a; 2021k) or Chief Fire and Rescue Service Advisor (2014). Instead, operational discretion was less likely to be used in the scenario in which it was licensed (the Discretion scenario) than in the scenario in which it was not licensed (the

Control scenario). Also, incident commanders were more likely to exhibit acute stress in the Discretion scenario than in the Control scenario, as measured by self-report and a blood-based measure of immunological function. Study 3b made use of THINCS to examine any differences in the use of command skills in the two scenarios. The principal difference between them was in situational awareness, which was less evident in the Discretion than the Control scenario. Finally, a thematic analysis of the answers given in a semi-structured interview was conducted to understand further the incident commanders who used operational discretion in a context-appropriate manner and those who used it inappropriately. This analysis revealed that those who used operational discretion appropriately (i.e., in the Discretion but not the Control scenario) judged that it was appropriate to accept a greater risk to firefighters in the context of the danger faced by the children, but not when there were no lives at risk. Those who used operational discretion inappropriately, judged the risk to firefighters to be appropriate in the control scenario, but not in the discretion scenario.

Overall, the research that led to the development and evaluation of the THINCS behavioural marker system has extended the use of such systems into the realm of the FRS. In accordance with other behavioural marker system research (e.g., Fletcher et al., 2003; Rutherford et al., 2015) the THINCS system was found to be reliable and valid. It also demonstrated its value as a research tool when investigating the application of operational discretion by incident commanders. A study that has extended the findings from laboratory-based research (e.g., Porcelli & Delgado, 2009; Schwabe et al., 2012; Starcke & Brand, 2012) that showed when under acute stress people tend to move away from deliberative decision making and increase their reliance on rules, into the incident command simulator-based environment of the UK FRS.

6.3 Limitations, future directions, and implications

The THINCS behavioural marker system is now used, under licence from Cardiff University, across different UK fire and rescue services. Studies 1, 2a and 2b provide a formal evaluation of the utility of THINCS. However, it is now being used for its intended purpose, to enable firefighters to develop their command skills through being given detailed feedback on their use of these skills. In fact, the system is already being used to determine readiness for promotion and revalidation of command skills. It could also be used to provide a database of skills across time for the UK FRS and to inform national operational learning in the UK FRS. A key implication of the introduction into the UK FRS of the THINCS system is that the extant suite of incident command qualifications and the incident command units in the FRS national occupational standards do not currently reflect a behavioural dimension. In general, they focus on knowledge, understanding and sets of performance criteria without explicit guidance about the command skills and the behaviours associated with them. Therefore, the THINCS system provides the UK FRS with an opportunity to update these to include a behavioural dimension linked to the command skills.

One potential development of the app would be to incorporate time-stamped videos of the incident commanders and voice to text facility for the evaluations. This has the potential to increase the efficacy of feedback and the usability of the app itself. One could also imagine linking the app to wearables in order to link the THINCS skills and feedback to indices of stress.

The use of simulated emergencies in studies 3a and 3b enables levels of reproducibility and experimental control that would be impossible in real emergencies: In particular, incidents requiring the use of operational discretion are relatively rare and the assessment of acute stress would be intrusive. However, simulations provide an incomplete representation of the variety and impacts of real emergency incidents on firefighter decision-making. For example, while the two simulated scenarios employed here had the predicted effects on measures of acute stress, they are unlikely to provide a complete representation of real emergency incidents, where the levels of acute stress are likely to be higher. A future study of complementary real-world incidents could clearly provide important converging evidence for conclusions based on those from simulations. It would also be beneficial to replicate the results reported in Chapters 4 and 5 in a broader range of scenarios; but there are obvious constraints on the availability of our participants to undertake research studies (and the diversity of the sample). One possibility would be to conduct a study in which the two scenarios were much more similar. For example, one could use high-rise fire scenarios in which operational discretion was required or not. Under these conditions, delivering the two scenarios to the same participants (i.e., a within-subjects procedure) would increase the power of the study with participants completing both scenarios. However, the disadvantage of a within-subjects design is that it would be subject to a number of confounds (e.g., transfer between the two scenarios). To avoid this, one would need to deliver them in a between-subjects procedure, which would increase the impact on the FRSs involved.

Nevertheless, the overall similarity between decision-making processes observed in real emergency incidents (Cohen-Hatton, Butler, & Honey, 2015) and a range of simulated ones (Cohen-Hatton & Honey, 2015), suggests that the results presented in Chapters 4 and 5 are very likely to generalise to real emergency incidents. Finally, it is possible that the nature of the scenarios has independent effects on the use of operational discretion and stress. However, accepting this possibility, leaves one without a ready explanation for why operational discretion was less likely to be used selectively in the scenario in which it is licensed than the scenario in which it was not.

The UK fire and rescue service guidance on operational decision making balances the efficiency of rules (SOPs) with the flexibility afforded by the (conditional) use of deliberation to respond to a wide variety of emergencies (i.e., operational discretion). My primary finding suggests that this balance is not reflected in operational decision making: operational discretion was more likely to be deployed when it was not licensed (in the Control scenario) than when it was licensed (in the Discretion scenario). The findings from Study 3b suggest this is linked to the incident commander's level of situational awareness which re-emphasises the important relationship between that and decision making (Endsley, 1995; Klein, 1993). In isolation, these results can be taken to suggest a need to reinforce the guidance and training. However, our secondary observation that the two scenarios were associated with different levels of acute stress suggests that this approach might be ineffective. To the extent that incidents licensing the use of operational discretion are likely to generate greater acute stress and this affects the use of SOPs and operational discretion. If one accepts the proposition that the use of operational discretion and deliberation should be licensed under unprecedented conditions (e.g., Grenfell Tower), then our results suggest a need for training to focus on generating effective situational awareness and decision making under stress perhaps coupled with personal resilience to mitigate the impact of acute stress on decision making (see Driskell, Johnston & Salas, 2001; Saunders, Driskell, Johnston & Salas, 1996). The clear prediction is that such training would increase the use of operational discretion when it is required.

The results of the survey of training provided across the UK FRS reported by Butler et al. (2020) are illuminating in this respect: All of the fire and rescue services that responded (27; approximately half of the UK fire and rescue services) delivered training in decision-making (25 involving both theoretical and practical components) and 25 delivered training in situational awareness (22 involving both theoretical and practical components). Yet, relatively few (14) provided training in any form of personal resilience (with only 9 providing practical training; cf. Sawhney, Jennings, Britt, & Sliter, 2018). There are clear grounds to augment training in personal resilience, directed at mitigating the effects of acute stress, given to first responders (cf. Conway & Waring, 2020); and to engender a culture in which different facets of incident command, including the effective use of operational discretion, are integrated and supported.

Finally, it is worth noting that THINCS was developed for incident commanders operating at a tactical level (i.e., located at the scene of the emergency). However, incident commanders at major incidents are supported by a multi-agency strategic co-ordinating group as well by tactical support officers, who are part of their command team. In order to support between-agency and within-agency working there is a need to integrate the training and evaluation of these teams, which is beyond the immediate scope of THINCS. To develop a behavioural marker system for command teams is one obvious extension to the current research.

6.4 Concluding comments

The research reported in this thesis was guided by a clear need to develop a system to evaluate the use of command skills by UK FRS incident commanders. It is clear that the use of non-technical (i.e., psychological) skills have an impact on the outcome of emergency incidents. The approach was informed by the precedent set by other high reliability industries and the methods used to do develop and evaluate behavioural marker systems (cf. Flin et al., 2008). This approach also reflected the need to meet the dual needs of developing a quantitative tool for the UK FRS (Chapters 2 and 3) with one that could also be used in a research context (Chapters 4 and 5). The research reported in Chapters 2 and 3 has already formed the basis of new training, evaluation and guidance for the UK FRS. The adoption of THINCS reflects the co-produced nature of the research that underpinned its development. The results reported in Chapters 4 and 5 suggest that stress interacts with processes that are at the heart of both the academic study of decision making and the guidance given to firefighters. Of course, further research is needed to reinforce the conclusions reached on the basis of the results from Chapters 4 and 5. In general terms, the research from Chapters 4 and 5 reinforce Simon's (1956) conclusion that decision makers are constrained by the information that they have, their cognitive limitations, and the time available to make a decision. In the context of the decisions made by firefighters, acute stress also influences their reliance on rules (SOPs) versus more deliberative processes.

References

Air Accident Investigation Bureau. (1990). *Report on the accident to Boeing 737-400G-OBME near Kegworth, Leicestershire on 8th January 1989. Aircraft Accident Report 4/90*. HMSO.

Almy, B., & Krueger, J. I. (2013). Game interrupted: The rationality of considering the future. *Judgment and decision making, 8*, 521.

Baker, S. R., & Williams, K. (2001). Short communication: Relation between social problem solving appraisals, work stress and psychological distress in male firefighters. *Stress and Health, 17*, 219–229.

Baron, J. (2000). *Thinking and deciding*. Cambridge University Press

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex,* 10, 295-307.

Bernoulli, D. (1738). Originally published in 1738; translated by Dr Louise Sommer. (January 1954)." Exposition of a New Theory on the Measurement of Risk. *Econometrica, 22*, 22-36.

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

Braun & Clarke. (2013). Teaching thematic analysis. *The Psychologist, 26,* 120-123.

Brunacini, A. (2002). Incident command functions. In R. Flin and K. Arbuthnot (Eds.), *Incident command: Tales from the hot seat.* (pp. 58-66).

Brunsden, V., Hill, R. & Maguire, K. (2014). Putting Fire & Rescue Service Stress Management into context: a UK informed perspective. *International Fire Service Journal of Leadership and Management, 7*, 27-39.

Bryman, A. (2012). *Social research methods* (4th ed). Oxford University Press.

Burke, E. (1997). Competence in command: recent R&D in the London Fire Brigade. *Journal of Managerial Psychology, 12,* 261-279.

Burke, L. A., & Hutchins, H. M. (2007). Training transfer: an integrative literature review. *Human Resource Development Review, 6*, 263–96.

Butler, P.C. (2016). *What are the non-technical skills of UK fire and rescue service incident commanders?* [Unpublished Masters dissertation] Cardiff University.

Butler, P. C., Honey, R. C., & Cohen-Hatton, S. R. (2020). Development of a behavioural marker system for incident command in the UK fire and rescue service: THINCS. *Cognition, Technology and Work, 22,* 1-12.

Butler, S. (1912). *The note-books of Samuel Butler.* Willian Brendon and Son Ltd.

Chief Fire Officers Association. (2015). *The future of incident command.* Chief Fire Officers Association.

Chief Fire and Rescue Service Advisor. (2014). *Fire and rescue authorities operational guidance: Generic risk assessment 3.2 fighting fires in high rise buildings.* The Stationary Office.

Cohen-Hatton, S.R., Butler, P.C., & Honey, R.C. (2015). An investigation of operational decision making in situ: incident command in the U.K. fire and rescue service. *Human Factors, 57,* 793-804.

Cohen-Hatton, S. R., & Honey, R. C. (2015). Goal-oriented training affects decision-making processes in virtual and simulated fire and rescue environments. *Journal of Experimental Psychology: Applied, 21,* 395–406.

Conway, R. R., & Waring, S. (2020). Facilitators and barriers to developing firefighter resilience. *Disaster Prevention and Management: An International Journal, 30,* 327-339.

Coombs, A. W., & Taylor, C. (1952). The effect of the perception of mild degrees of threat on performance. *The Journal of Abnormal and Social Psychology, 47p,* 420-424.

Coultas, C. W., Grossman, R. & Salas, E. (2012). Design, delivery, evaluation, and transfer of training systems. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (4th ed., pp. 490–533). John Wiley & Sons.

Cox, T. (1978). *Stress*. Macmillan.

Crichton, M. T., & Flin, R. (2004). Identifying and training non-technical skills of nuclear emergency response teams. *Annals Nuclear Energy, 31,* 1317–1330.

Crichton, M. T., Lauche, K., and Flin, R. (2005). Incident command skills in the management of an oil industry drilling incident: a case study. *Journal of Contingencies and Crisis Management*, *13*, 116–128.

Crichton, M. T., Moffat, S., & Crichton, L. (2017). Developing a team behavioural Marker framework using observations of simulator-based exercises to improve team effectiveness: A drilling team case study. *Simulation Gaming, 48,* 299-313.

Crichton, M. T., Ramsay, C. G., & Kelly, T. (2009). Enhancing organizational resilience through emergency planning: learnings from cross-sectoral lessons. *Journal of Contingencies and Crisis Management*, *17*, 24-37

da Conceição, V. P., Basso, J. C., Lopes, F. C., & Dahlman, J. (2017). Development of a behavioural marker system for rating cadet's non-technical skills. *TransNav the International Journal on Marine Navigation and Safety of Sea Transportation, 11,* 255-262.

DCLG. (2013). *Fire and rescue authorities health, safety and welfare framework for the operational environment.* GOV.UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/209362/HSFrameworkJunecombined.pdf

DCLG. (2011). *The economic cost of fire: Estimates for 2008. Fire research report 3/2011*. Department for Communities and Local Government.

De Vellis, R. F. (2005). Inter-Rater Reliability. In K. Kempf-Leonard (Ed). *Encyclopedia of Social Measurement* (pp 317-322). Elsevier.

Devitt, K., & Holford, S. (2010). The development of resource management and leadership behavioural markers for the Merchant Navy. Solent University. https://ssudl.solent.ac.uk/id/eprint/1426/

Driskell, J. E., Johnston, J. H., & Salas, E. (2001). Does stress training generalize to novel settings? *Human Factors*, *43*, 99–110.

Driskell, T., Salas, E., & Driskell, J. E. (2018). Teams in extreme environments: Alteration in team development and teamwork. *Human Resource Management Review, 28,* 434-449. http://dx.doi.org/10.1016/j.hrmr.2017.01.002

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*(3), 183–201. https://doi.org/10.1037/h0047707

Edmondson, A. C. (2003). Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies*, *40*, 1419–1452.

Endsley, M. R. (1989). *Final report: Situation awareness in an advanced strategic mission (NOR DOC 89-32).* Northrop Corporation.

Endsley, M. R. (1995). Towards a theory of situational awareness in dynamic systems. *Human Factors, 37,* 32-64.

Endsley, M. R. (1997). Supporting situation awareness in aviation systems. *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, 5*, 4177-4181. https://doi.org/10.1109/ICSMC.1997.637352

Endsley M. R. (2021). A systematic review and meta-analysis of direct, objective measures of SA: A comparison of SAGAT and SPAM. *Human Factors, 63,* 124-150.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*, 709–724.

Fellows, L. K. (2004). The cognitive neuroscience of human decision making: A review and conceptual framework. *Behavioural Cognitive Neuroscience Review, 3*, 159-172.

Fennell, D. (1988). *Investigation into the Kings Cross fire.* HMSO.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed). Sage.

Fire Brigades Union. (2008). *In the line of duty: Firefighter deaths since 1978.* Fire Brigades Union. https://www.fbu.org.uk/publication/line-duty-firefighter-deaths-uk-1978

Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N., & Patey, R. (2003). Anaesthetists' non-technical skills (ANTS): evaluation of a behavioural marker system. *British Journal of Anaesthesia, 90,* 580-588.

G. Fletcher, R. Flin, P. McGeorge, R. Glavin, N. Maran, R. Patey. (2003). Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system, *British Journal of Anaesthesia, 90*, 580–588, https://doi.org/10.1093/bja/aeg112

Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N., & Patey, R. (2004). Rating non-technical skills: Developing a behavioural maker system for use in anaesthesia. *Cognition, Technology and Work, 6*, 165–171.

Flin, R., & Martin, L. (2001). Behavioural markers for crew resource management: A review of current practice. *The International Journal of Aviation Psychology*, *11*, 95–118.

Flin, R., Martin, L., Goeters, K-M., Hörmann, H-J., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety, 3,* 95–117.

Flin, R., O'Connor, P., & Crichton, M. (2008). *Safety at the sharp end a guide to non-technical skills*. Ashgate.

Flin, R., O'Connor, P., & Mearns, K. (2002). Crew resource management: improving team work in high reliability industries. *Team Performance Management: An International Journal, 8,* 68-78.

Flin, R., Youngson, G., & Yule, S. (2016). *Enhancing surgical performance A primer in non-technical skills.* CRC Press.

Friedman, H. H., & Amoo, T. (1999). Rating the rating scales. *Journal of Marketing Management, 9*, 114-123

Friedman, H. H., & Leefer, J. R. (1981), Label versus position in rating scales, *Journal of the Academy of Marketing Science, 9*, 88-92.

Gaba, D., Howard, S., Flanagan, B., Smith, B., Fish, K., & Botney, R. (1998). Assessment of clinical performance during simulated crises using both technical and behavioural ratings. *Anaesthesiology, 89,* 8-18.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.

Goldstein, W. M., & Hogarth, R. M. (1997). Judgement and decision research: Some historical context. In W. M. Goldstein & R. M. Hogarth (Eds.), *Research on judgement and decision making: Currents, connections, and controversies (pp. 3-68). Cambridge University Press.*

Greater Manchester Fire and Rescue Service (GMFRS). (2016). *Fatal accident investigation report. Paul's hair and beauty world, Oldham Street, Manchester. July 13 2013.*

Greater Manchester Fire and Rescue Authority. https://www.manchesterfire.gov.uk/media/1835/fatal-accident-investigation-report-oldham-street-july-2013-web.pdf

Grenfell Tower Inquiry. (2019). *Phase 1 report. Report of the public inquiry into the fire at Grenfell Tower on 14 June 2017.* (Vol. 2). (OGL Publication HC 49 II). Grenfell Tower Inquiry. https://assets.grenfelltowerinquiry.org.uk/GTI%20-%20Phase%201%20full%20report%20-%20volume%202.pdf

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In: N. K. Denzin, & Y. S. Lincoln (Eds). *Handbook of qualitative research* (pp. 105–117). Sage

Hallet, H. C. (2011). *Coroner's inquests into the London bombings of July 5th 2005.* National Archives: http://webarchive.nationalarchives.gov.uk/20120216072438/http://7julyinquests.independent.gov.uk/docs/orders/rule43-report.pdf

Hamlet, O., Irwin, A., Flin, R., & Sedlar, N. (2020a). HELINOTS: Non-technical skills for offshore transport pilots. Aberdeen University. https://research.abdn.ac.uk/wp-content/uploads/sites/14/2020/07/HeliNOTS-O-Handbook.pdf

Hamlet, O., Irwin, A., Flin, R., & Sedlar, N. (2020b). HELINOTS: Non-technical skills for helicopter search and rescue. Aberdeen University. https://research.abdn.ac.uk/wp-content/uploads/sites/14/2021/02/HeliNOTS-SAR-Handbook.pdf

Haslam, C., & Mallon, K. (2001). A preliminary investigation of post-traumatic stress symptoms among firefighters. *Work & Stress, 17,* 277-285.

Hastie, R. (2001). Problems for judgment and decision making. *Annual Review of Psychology*, 52, 653-683.

Hastie, R. & Dawes, R. M. (2010). *Rational choice in an uncertain world. The psychology of judgement and decision making* (2nd ed). Sage.

Health Commission. (2009). *Investigation into Mid Staffordshire NHS foundation trust.* British Library. https://www.bl.uk/collection-items/investigation-into-mid-staffordshire-nhs-foundation-trust

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *International Journal of Aviation Psychology, 9,* 19-32.

HM Government. (2008). *Fire and rescue service manual volume 2 fire service operations incident command* (3rd ed). Assets Publishing Service. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/7643/incidentcommand.pdf

Holt, R., Boehm-Davis, D., & Beaubien, M. (2001). Evaluating resource management in training. E. Salas, C. Bowers & E. Edens (Eds.), *Applying resource management in organizations: A guide for training professionals* (pp. 165-187). Lawrence Erlbaum Assoc.

HSE. (1999). *HSG48: Reducing error and influencing behaviour.* Health and Safety Executive. https://www.hse.gov.uk/pubns/books/hsg48.htm

HSE. (2010a) *Striking the balance between operational and health and safety duties in the Fire and Rescue Service.* Health and Safety Executive. http://www.hse.gov.uk/services/fire/duties.pdf

HSE (2010b) *Management of health and safety in the GB fire and rescue service consolidated report on the 8 inspections completed by HSE in 2009/10.* HSE. https://www.hse.gov.uk/services/fire/management.pdf

Hutchby, I., & Wooffitt, R. (1998). *Conversation analysis: Principles, practices and applications.* Polity Press.

International Association of Oil and Gas Producers. (2018). *Introducing behavioural markers of non-technical skills in oil and gas operations* (503). https://www.hpog.org/assets/documents/IOGP-503.pdf

Janis, I. L., & Mann, L. (1977). Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment. Free Press.

JESIP. (2016). *The interoperability framework (2nd ed)*. JESIP. https://www.jesip.org.uk/uploads/media/pdf/Joint%20Doctrine/JESIP_Joint_Doctrine_Document.pdf

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697–720.

Kahneman, D. (2011). *Thinking fast and thinking slow.* Penguin.

Kahneman, D., & Tversky, A (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 264–291.

Kanki, B., Anca, J., & Chidester, T. R. (2019) (Eds.), *Crew Resource Management.* San Diego: Academic Press

Kassam, K. S., Koslov, K., & Mendes, W. B. (2009). Decisions under distress: Stress profiles influence anchoring and adjustment. *Psychological Science*, *20*, 1394-1399.

Kiecolt-Glaser, J. K., Marucha, P. T., Malarkey, W. B., Mercado, A. M., & Glaser, R. (1995). Sowing of wound healing by psychological stress. *The Lancet, 346,* 1194-1196.

Kirkham, F. (2013). Coroner's Inquest into the Lakanal House Fire of 3rd July 2009 Rule 43 Letter. Lambeth Council website: http://www.lambeth.gov.uk/sites/default/files/ec-letter-to-london-fire-brigade-pursuant-to-rule43-28March2013.pdf

Klaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors, 38,* 220-231.

Klampfer, B., Flin, R., Helmreich, R. L., Häuser, R., Sexton, B., Fletcher, G., Field, P., Staender, S., Lauche, K., Dieckmann, A., & Amacger, A. (2001). *Enhancing Performance in High Risk Environments: Recommendations for the Use of Behavioural Markers.* Daimler-Benz Stiftung.

Klein, G. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. Klein, J. Orasanu, R. Calderwood and C. Zsambok (Eds.), *Decision making in action* (pp. 138-147). Ablex.

Klein, G. (1998). *Sources of power: How people make decisions.* MIT Press.

Klein G. A. (2008) Naturalistic decision making. *Human Factors, 50*, 456–460

Klein, G. A. (2009). *Streetlights and shadows: Searching for the keys to adaptive decision making.* MIT Press.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics*, *19*, 462–472.

Klein, G., Shneiderman, B., Hoffman, R.R., & Ford, K.M. (2017). Why expertise matters: A response to the challenges. *IEEE Intelligent Systems, 32,* 67-73.

Kodate, N., Ross, A. J., Anderson, J. E., Flin, R. (2012). Non-technical skills (NTS) for enhancing patient safety: Achievements and future directions. *Japanese Journal of Quality and Safety in Healthcare, 7*, 360–370.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brian Sciences, 19,* 1-17.

Koolhaas, J. M., Bartolomucci, A., Buwalda, B., de Boer, S. F., Flugge, G., Korte, S. M., Meerlo, P., Murison, R., Olivier, B., Palanza, P., Richter-Levin, G., Sgoifo, A., Steimer, T., Stiedl, O., van Dijk, G., Wohr, M., Fuchs, E.. (2011). Stress revisited: a critical evaluation of the stress concept. *Neuroscience Biobehavioral Reviews, 35*, 1291–1301.

Kudielka, B. M., Hellhammer, D. H., Wust, S., 2009. Why do we respond so differentially? Reviewing determinants of human salivary cortisol responses to challenge. *Psychoneuroendocrinology, 34*, 2–18.

LaPorte, T. R., & Consolini, P. M. (1991). Working in Practice but Not in Theory: Theoretical Challenges of "High-Reliability Organizations". *Journal of Public Administration Research and Theory, 1,* 19-48.

Lauber, J. (1984). "Resource management in the cockpit". *Air Line Pilot, 53,* 20-23.

Lazarus, R., & Folkman, S. (1984). *Stress, Appraisal, and Coping*. Springer.

Lekka, C. (2011). *High reliability organisations a review of the literature*. HSE Books

Lipshitz, R., & Bar-Ilan, O. (1996). How problems are solved: Reconsidering the phase theorem. *Organizational Behavior and Human Decision Processes, 65*, 48-60.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Focus article: Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14,* 331-352.

List of British firefighters killed in the line of duty. (19 July 2021). In *Wikipedia*. https://en.wikipedia.org/wiki/List_of_British_firefighters_killed_in_the_line_of_duty

London Fire Brigade. (2004). Decision making model. Grenfell Tower Inquiry. https://www.grenfelltowerinquiry.org.uk/evidence/lfb-policy-documents

Louie, M. A. & Carley, K. M. (2008). Balancing criticisms: Validating multi-agent models of social systems. *Simulation Modelling Practice and Theory, 16,* 242-256.

Marine Casualty Investigative Body. (2012). *Cruise ship Costa Concordia marine casualty on January 13, 2012 report on the safety technical investigation*. Spanish Ministry of Infrastructures and Transport. https://www.mitma.gob.es/recursos_mfom/2012costaconcordia.pdf

Mariotti, A. (2015). The effects of chronic stress on health: new insights into the molecular mechanisms of brain-body communication. *Future Science OA, 1,* FSO23.

Mark, G. M., & Smith, A. P. (2008). Stress models: A review and suggested new direction. In: *Occupational Health Psychology: European Perspectives on research, education and practice*, *3*, 111-144.

Mason, J.W., 1968. A review of psychoendocrine research on the sympathetic‐adrenal medullary system. *Psychosomatic Medicine, 30*, 631‐653.

McCulloch, P., Mishra, A., Handa, A., Dale, T., Hirst, G., & Catchpole, K. (2009). The effects of aviation-style bob-technical skills training on technical performance and outcome in the operating theatre. *BMJ Quality and Safety in Health Care, 18,* 109-115.

McLennan, J., Strickland, R., Omodei, M., & Suss, J. (2014). Stress and wildland firefighter safety-related decisions and actions. In C. Owen (Ed.), *Human factors challenges in emergency management: Enhancing individual and team performance in fire and emergency management* (pp. 19–33). CRC Press.

Mitchell, L., Flin, R., Yule, S., Mitchell, J., Coutts, K., and Youngson, G. (2013). Development of a behavioural marker system for scrub practitioners' non-technical skills (SPLINTS) system. *Journal of Evaluation in Clinical Practice, 19,* 317–323.

Mosier, K. L., & Fischer, U. M. (2010). Judgment and Decision Making by Individuals and Teams: Issues, Models, and Applications. *Reviews of Human Factors and Ergonomics, 6,* 198–256. https://doi.org/10.1518/155723410X12849346788822

National Audit Office. (2018). *Investigation into government-funded inquires* (HC 836). National Audit Office. https://www.nao.org.uk/wp-content/uploads/2018/05/Investigation-into-government-funded-inquiries.pdf

N.O.G. (2021a). Operational discretion. Fire Central Programme Office. https://www.ukfrs.com/guidance/search/operational-discretion

N.O.G. (2021b). Situational awareness. Fire Central Programme Office. https://www.ukfrs.com/guidance/search/situational-awareness

N.O.G. (2021c). Guidance. Fire Central Programme Office. https://www.ukfrs.com/guidance

N.O.G. (2021d). Command skills and the role of the incident commander. Fire Central Programme Office. https://www.ukfrs.com/guidance/incident-command?bundle=section&id=31283&parent=12550

N.O.G. (2021e). Scalable command arrangements. Fire Central Programme Office. https://www.ukfrs.com/foundation-knowledge/incident-command-knowledge-skills-and-competence?bundle=section&id=17019&parent=17010

N.O.G. (2021f). Levels of command. Fire Central Programme Office. https://www.ukfrs.com/guidance/search/levels-command

N.O.G. (2021g). Incident command: Knowledge, skills and competence. Fire Central Programme Office. https://www.ukfrs.com/foundation-knowledge/incident-command-knowledge-skills-and-competence

N.O.G. (2021h). Firefighter maxim. Fire Central Programme Office. https://www.ukfrs.com/promos/17037

N.O.G. (2021i). Complementing incident command experience: A guide for fire and rescue. Fire Central Programme Office. https://www.ukfrs.com/foundation-knowledge/complementing-incident-command-experience-guide-fire-and-rescue-services?bundle=section&id=32190

N.O.G. (2021j). Control measure: Decision making. Fire Central Programme Office. https://www.ukfrs.com/guidance/search/decision-making

N.O.G. (2021k). Fires in tall buildings. Fire Central Programme Office. https://www.ukfrs.com/guidance/search/fires-tall-buildings

N.O.G. (2021l). Behavioural marker systems. Fire Central Programme Office. https://www.ukfrs.com/guidance/search/behavioural-marker-systems

Nunnally, J. C. (1967). *Psychometric Theory.* McGraw-Hill.

O'Connor, P., & Long, W. M. (2011). The development of a prototype behavioral marker system for US Navy officers of the deck. *Safety Science, 49,* 1381–1387.

O'Connor, P., O'Dea, A., Flin, R., & Belton, S. (2008). Identifying the team skills required by nuclear power plant operations personnel. *International Journal of Industrial Ergonomics, 38,* 1028-1037.

Office for National Statistics. (2021). *Average weekly earnings in Great Britain: July 2021***.** Office for National Statistics**.** https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/averageweeklyearningsingreatbritain/july2021

Okray, R., & Lubnau II, T. (2004). *Crew resource management for the fire service.* PennwWell.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 3–21). Ablex.

Peters, A., McEwen, B. S., & Friston, K. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in Neurobiology*, *156*, 164-188.

Pollack, S., Friedman, H. H. & Presby, L. (1990, November 11-14). *Two salient factors in the construction of rating scales: Strength and direction of anchoring adjectives* [Conference paper]. International Conference of Measurement Errors in Surveys. Tucson, Arizona.

Porcelli, A. J., & Delgado, M. R. (2009). Acute stress modulates risk taking in financial decision making. *Psychological Science*, *20*, 278-283.

Porcelli, A. J., & Delgado, M. R. (2017). Stress and decision making: Effects on valuation, learning, and risk-taking. *Current Opinion in Behavioral Sciences*, *14*, 33-39.

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour.* Sage.

Qualtrics. (2019). *Qualtrics* (Version December 2019) [Computer Software]. Qualtrics https://www.qualtrics.com/uk/

QSR International. (2018). *NVivo* (Version 12) [Computer software]. NVivo https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Railway Safety and Standards Board. (2012). *Operations and Management - Non-technical skills for rail: A list of skills and behavioural markers for drivers, with guidance notes (2012)* (T869 Report). Spark. https://www.sparkrail.org/Lists/Records/DispForm.aspx?ID=10025

Rafferty, L. A., Stanton, N. A., & Walker, G. H. (2013). Great expectations: A thematic analysis of situation awareness in fratricide. *Safety Science*, *56*, 63–71. https://doi.org/10.1016/j.ssci.2012.03.020

Raykov, T., & Marcoulides, G. A. (2010). *Introduction to Psychometric Theory.* Routledge.

Reason, J. (2016). *Organizational accidents revisited.* Ashgate.

Roberts, R., Flin, R., & Cleland, J. (2015). Staying in the zone: Offshore drillers' situation awareness. *Human Factors, 57,* 573-590.

Roberts, K., & Rousseau, D. M. (1989). Research in nearly failure-free, high reliability organisations: Having the bubble. *IEEE Transactions on Engineering Management*, *36*, 132- 139.

Rutherford, J. S., Flin, R., Irwin, A., & McFadyen, A. K. (2015), Evaluation of the prototype anaesthetic non-technical skills for anaesthetic practitioners (ANTS-AP): a behavioural marker system to assess the non-technical skills used by staff assisting the anaesthetist. *Anaesthesia, 70,* 907-914.

Salas, E., Driskell, J. E., & Hughes, S. (1996). Introduction: The study of stress and human performance. In J. E. Driskell & E. Salas (Eds.), *Stress and human performance* (pp. 1-45). Erlbaum.

Salas, E. Wilson, K. A., Burke, C. S. & Wightman, D. C. (2006). Does crew resource management training work? An update, an extension, and some critical needs, *Human Factors, 48,* 392-412.

Salmon, P. M., Stanton, N. A., Walker, G. H., & Jenkins, D. P. (2009). *Distributed situation awareness.* CRC Press.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology, 1,* 45-57.

Saunders, T., Driskell, J. E., Johnston, J. H., & Salas, E. (1996). The effect of stress inoculation training on anxiety and performance. *Journal of Occupational Health Psychology, 1*(2), 170–186. https://doi.org/10.1037/1076-8998.1.2.170

Savage, L. (1954). *The Foundations of Statistics*. John Wiley & Sons, Inc.

Sawhney, G., Jennings, K. S., Britt, T. W., & Sliter, M. T. (2018). Occupational stress and mental health symptoms: Examining the moderating effect of work recovery strategies in firefighters. *Journal of Occupational Health Psychology*, *23*, 443-456.

Schuermann, V. & Marquardt, N. (2016). Adaptation of crew resource management training in high-risk industries, *International Journal of Safety and Security Engineering, 6,* 341–350.

Schwabe, L., Tegenthoff, M., Hoffken, O., & Wolf, O. T. (2012). Simultaneous glucocorticoid and noradrenergic activity disrupts the neural basis of goal-directed action in the human brain. *Journal of Neuroscience,* 32, 10146-10155.

Selye, H. (1956). *The stress of life.* McGraw-Hill.

Shelton-Rayner, G. K., Macdonald, D. W., Chandler, S., Robertson, D., & Mian, R. (2010). Leukocyte reactivity as an objective means of quantifying mental loading during ergonomic evaluation. *Cellular Immunology*, *263*, 22-30.

Shiv, B., Loewenstein, G., Bechara, A., Damasio, H., & Damasio, A. R. (2005). Investment behavior and the negative side of emotion. *Psychological Review, 16,* 435-439.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63, 129.

Simon, H. A. (1978). Rationality as a process and product of thought. *American Economic Association, 68,* 1-16.

Skills for Justice. (2013a). SFJ EFSM1 Provide strategic advice and support to resolve operational incidents. National Occupational Standards: https://www.ukstandards.org.uk/PublishedNos-old/SFJEFSM1.pdf#search=SFJ%20EFSM1

Skills for Justice. (2013b). SFJ EFSM2 Lead, monitor and support people to resolve operational incidents. National Occupational Standards. https://www.ukstandards.org.uk/PublishedNos-old/SFJEFSM2.pdf#search=SFJ%20EFSM2

Skills for Justice. (2014). SFJ FRS WM7 Lead and support people to resolve operational incidents. National Occupational Standards: https://www.ukstandards.org.uk/PublishedNos-old/SFJFRSWM7.pdf#search=SFJ%20WM7

Skills for Justice. (n.d.). Qualifications handbook incident command in fire and rescue services. Skills for Justice Awards: http://www.sfjawards.com/wp-content/uploads/2013/10/Qualifications-Handbook-FRS-Incident-Command-V3.0.pdf

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Smith A. P., Smith H. N. (2017). Workload, fatigue and performance in the rail industry. In: Longo L., Leva M. (Eds.), *Human Mental Workload: Models and Applications*. H-WORKLOAD 2017. Communications in Computer and Information Science, vol 726. Springer.

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis theory, method and research.* Sage.

Staal, M. (2004). Stress, cognition, and human performance: A literature review and conceptual framework. National Aeronautical Space Agency. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20060017835.pdf

Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. *Theoretical issues in Ergonomics Science, 1,* 315-331.

Starke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience & Biobehavioral Reviews*, *36*, 1228-1248.

Stokes, A.F., & Kite, K. (2001). On grasping a nettle and becoming emotional. In P. A. Hancock, & P. A. Desmond (Eds.), *Stress, workload, and fatigue*. Erlbaum

Thomas, M. J.W. (2018). *Training and assessing non-technical skills a practical guide.* CRC Press.

Todd, P.M., and Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, *16*, 167-171.

Torrie, S. (2012). *A report to Scottish Ministers the 2008 Galston mine incident.* APS Group Scotland.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review, 79,* 281-299.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Urbina, S. (2004). *Essentials for psychological testing.* John Wiley & Sons.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton University Press.

Watterson, A. (2015). *Fire-fighter fatalities at fires in the UK: 2004–2013: Voices from the fireground.* Fire Brigades Union. https://www.fbu.org.uk/publication/briefing-firefighter-fatalities-fires-uk-2004-2013-voices-fireground

Webster Marketon, J. I., & Glaser, R. (2008). Stress hormones and immune function. *Cellular Immunology, 252,* 16-26.

Weick, K. (1990). The vulnerable system: an analysis of the Tenerife air disaster. *Journal of Management*, *16*, 571–593.

Wickens, C. D., Lui, Y., Becker, S. E. G., & Lee, D. (2004). *An introduction to human factors engineering* (2nd ed). Prentice Hall

Wilkinson, B., Cohen-Hatton, S. R., & Honey, R. C. (2019). Decision-making in multi-agency groups at simulated major incident emergencies: In-situ analysis of adherence to UK doctrine. *Journal of Contingencies and Crisis Management, 27,* 306-316.

Williams, G., Pendlebury, H. & Smith, A. P. (2017). Stress and well-being of nurses: An investigation using the Demands-Resources-Individual Effects (DRIVE) model and Well-being Process Questionnaire (WPQ). *Open Access Journal of depression and Anxiety*, *1*, 1-8.

Yates, J. F., Veinott, E. S., & Patalano, A. L. (2003). Hard decisions, bad decisions: On decision quality and decision aiding. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 13-63). Cambridge University Press.

Youngson, G. G. (2016). Intraoperative performance, non-technical skills and surgical safety. In: R. Flin, G. G. Youngson, & S. Yule (Eds.), *Enhancing surgical performance: A primer in non-technical skills* (pp. 3–16). CRC Press.

Yule, S., Flin, R., Paterson-Brown, S, Maran, N., & Rowley, D. (2006). Development of a rating system for surgeons' non-technical skills. *Medical Education, 40*, 1098–1104.

Zsambok, C. E. (1997). Naturalistic decision making: Where are we now? In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 3-16). Lawrence Erlbaum & Associates.

Zsambok, C. E., & Klein, G. (2014). *Naturalistic Decision Making*. Hove.

Appendices

Appendix 1: THINCS Rater Questionnaire

Rater questionnaire

Comprehensiveness of the System		**Yes**		**No**	
Were there command skills or sub-skills that were observed but that were omitted from the system?		☐		☐	
Comments:					
Were there command skills or sub-skills that were included in the system but that were unnecessary?		☐		☐	
Comments:					

*5 = Very Easy; 4 = Easy; 3 = Average; 2 = Difficult; 1 = Very Difficult

Observability of Command Skills*	**5**	**4**	**3**	**2**	**1**
How easy was it to link behaviours to the relevant command skills?	☐	☐	☐	☐	☐
Comments:					
How easy was it to link behaviours to the sub-skills?	☐	☐	☐	☐	☐
Comments:					

Utility of the System		**Yes**		**No**	
Were the command skills and sub-skills clearly defined?		☐		☐	
Comments:					
Would the wording used to describe the command skills and sub-skills be meaningful in a UK FRS context?		☐		☐	
Comments:					

Were the ‘good practice’ behavioural markers useful?	☐	☐
Comments:		
Were the ‘poor practice’ behavioural markers useful?	☐	☐
Comments:		

Utility of the Rating Scale	**Yes**	**No**
Were the descriptors clearly defined?	☐	☐
Comments:		
Was the five point rating scale (0 – 4), plus a ‘not observed’ category an appropriate rating scale?	☐	☐
Comments:		
Was the rating scale sufficiently flexible to rate the performance levels observed?	☐	☐
Comments:		

*5 = Very Easy; 4 = Easy; 3 = Average; 2 = Difficult; 1 = Very Difficult

	5	**4**	**3**	**2**	**1**
How easy was it to use part one of the observation pro forma?	☐	☐	☐	☐	☐
Comments:					

	Yes	**No**
Was the format of part one of the observation pro forma suitable for recording observations?	☐	☐
Comments:		
Was the format of part two of the observation pro forma (rating and feedback) suitable for collating feedback?	☐	☐

Comments:					
Was the format of the incident commander feedback document suitable?		☐		☐	

*5 = Very good; 4 = Good; 3 = Acceptable; 2 = Poor; 1 = Very Poor

Suitability of the Video Clips	**1**	**2**	**3**	**4**	**5**
Rate the quality of the video clips*?	☐	☐	☐	☐	☐
Comments:					
Rate the content of the video clips*?	☐	☐	☐	☐	☐
Comments:					

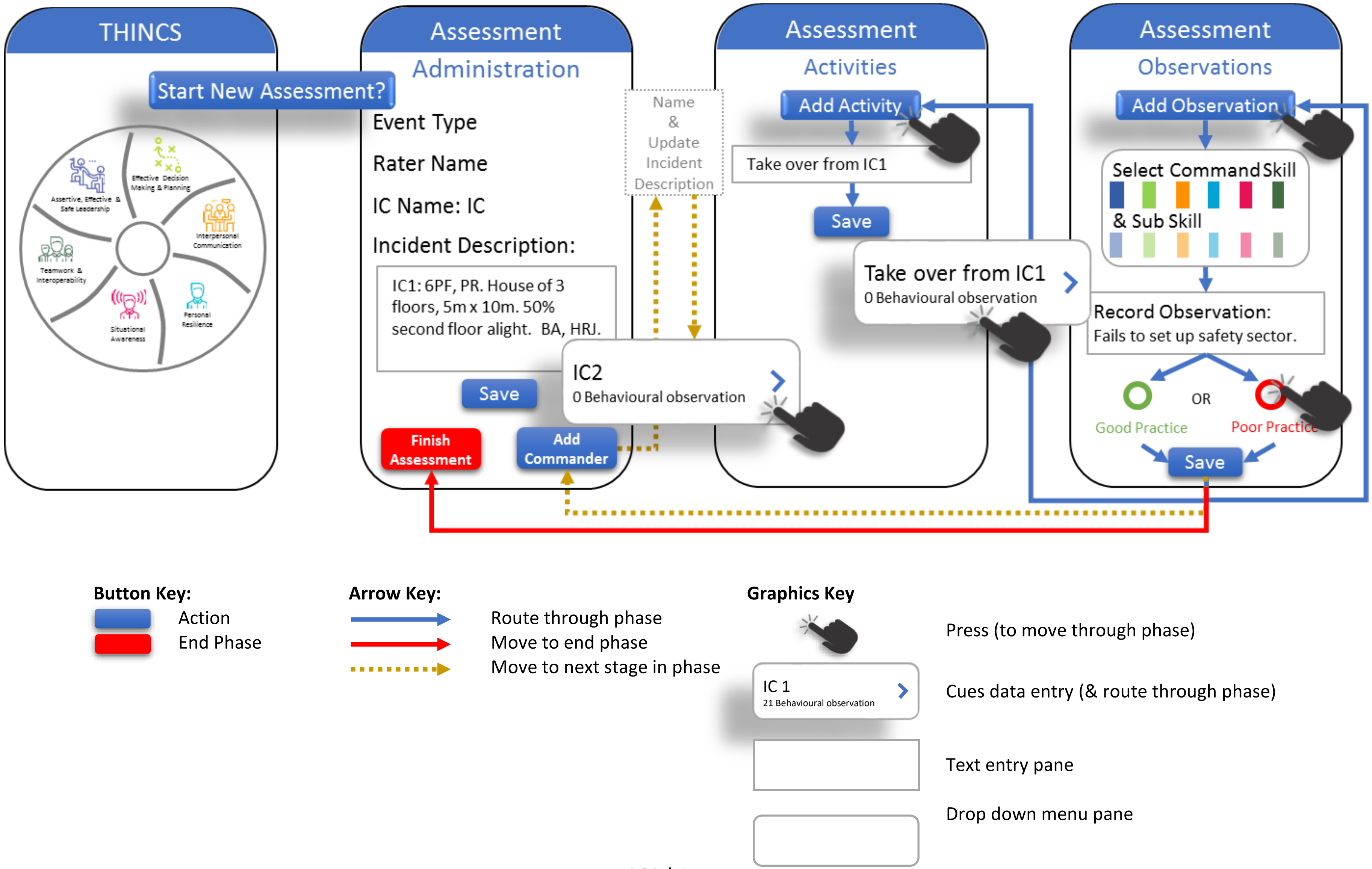
		Yes		**No**	
Were the video clips sufficiently long enough?		☐		☐	
Comments:					

***1 = Best; 2 = Satisfactory; 3 = Worst

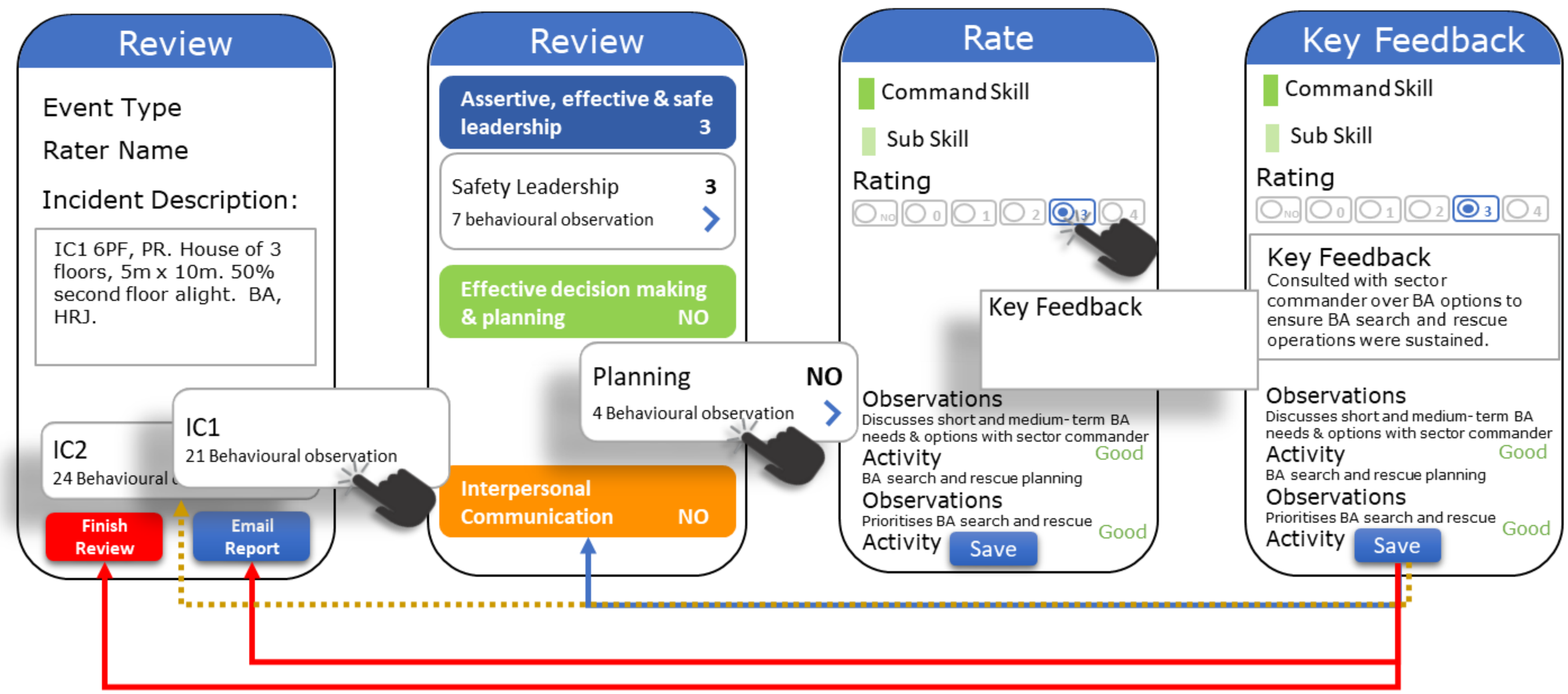
	Real Incident	**Practical Exercise**	**Compute Suite Simulation**
Rank the type of video clips used***?			
Comments:			

Appendix 2: THINCS App Schematics

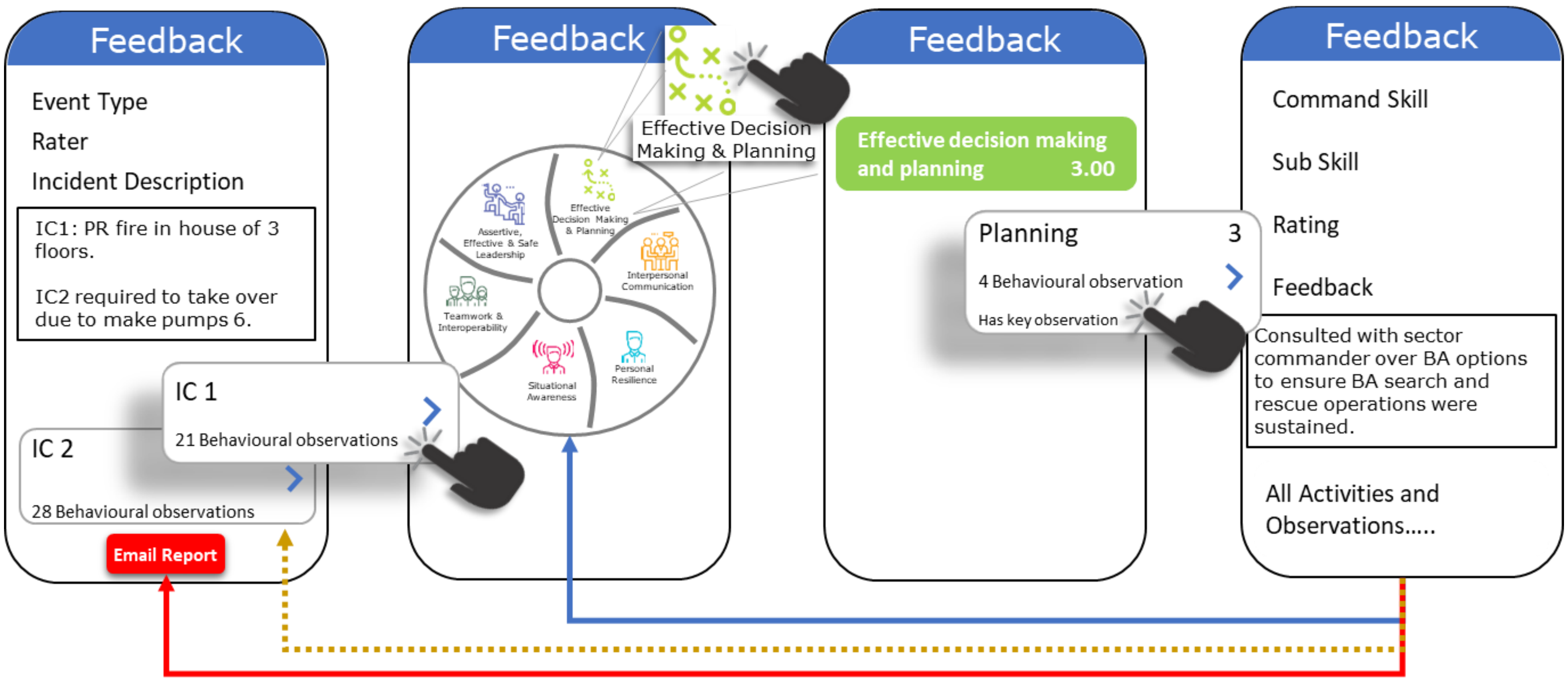
Administration and Observations Schematic



Review, Rate and Identify Key Feedback Schematic



Feedback Provision Schematic



Appendix 3: THINCS Rater Course Syllabus

THINCS Rater Course
Course Content
Module 1
• Human factors and safety
○ Definition of human factors
○ Importance of human factors to high reliability organisations
○ Relationship between human error and:
▪ Accident causation
▪ Command skills
○ Two views of human error
• Error:
○ Organisational Error Model
○ Human error
▪ Definition
▪ Human characteristics towards making errors
▪ Types of error
▪ Classifications of human error
▪ Violations
○ Error types
• Human performance:
○ Relationship between human variability, context and error
○ Human information processing
○ Three levels of human performance and associated error types
○ Effects of expertise, stress and fatigue on performance
○ Effects of human limitations:
▪ Perception
▪ Attention
▪ Memory
• Human fallibility:
○ Associative thinking
○ Priming
○ Heuristics
○ Cognitive ease
▪ Representativeness
▪ Availability
▪ Anchoring and adjustment
○ Cognitive biases
▪ Human susceptibility toward bias
▪ Categories of cognitive bias
Module 2
• Performance rating:
○ Subjective and objective judgements
○ Attribution
○ Free modulus estimation
○ Inter-rater agreement and inter-rater reliability
○ Characteristics of a reliable rating system:

        - Reliability
        - Validity
        - Utility
- THINCS behavioural marker system:
    - Rating Scale
    - Documents
    - Methodology
    - Administration
    - Guided practise rating
    - Rating exercises

Module 3

- THINCS behavioural marker system:
    - Overview of THINCS App
- Practise sessions
    - THINCS assessment
    - Rating review
    - Rating consensus
- Assessments:
    - Knowledge
    - Practical

Appendix 4: THINCS Evaluation Project Focus Group Questions

Questions:

General IC Assessment

1. What do you want out of incident commander assessments (summative and formative)?

2. What is important when assessing an incident commander?

3. How do you assess these important factors?

    a) Is it the same for all levels of command?

Command Skills

4. What do you think of non-technical skills?

5. What is important about non-technical skills?

6. What do you think is the relationship between non-technical and technical command skills?

    a) Do they complement/clash with each other?

    b) How do they complement/clash with each other?

7. What is your impression of the command skills that underpin THINCS?

    a) What specific command skills impressed/disappointed you?

    b) Are there any command skills missing/unnecessary?

8. Do you feel the command skills and sub skills are well defined?

THINCS Behavioural Marker System

General

9. What do you think of the THINCS behavioural marker system?

    a) The methodology of Observe, review, rate and feedback?

    b) The rating system?

        i) Is the rating scale sufficiently flexible to rate performance?

        ii) Do you think the descriptors were suitable?

10. How easy is it to link behaviours to the relevant command skills/sub skills?

11. Are the exemplar behaviours of good and poor practice useful?

THINCS Paper-based System

12. What do you feel about the paper-based documents?

    a) The format of the observation/feedback pro forma?

b) How would you describe the feedback process? (only ask ICs)

c) The structure of the database?

13. How easy is it to use the observation/feedback document?

14. How could the paper-based documents be improved?

THINCS Mobile App

15. What do you think of the mobile app?

a) How easy is it to install?

b) How intuitive is it to navigate through the methodology?

c) How easy is it to interpret the information and symbols on the screens?

16. What impressed/disappointed you about the app?

a) How would you describe the feedback process? (only ask ICs)

17. How could the app be improved?

18. What are the advantages/disadvantages of using the app compared to the paper-based system? (only ask if participants have used both)

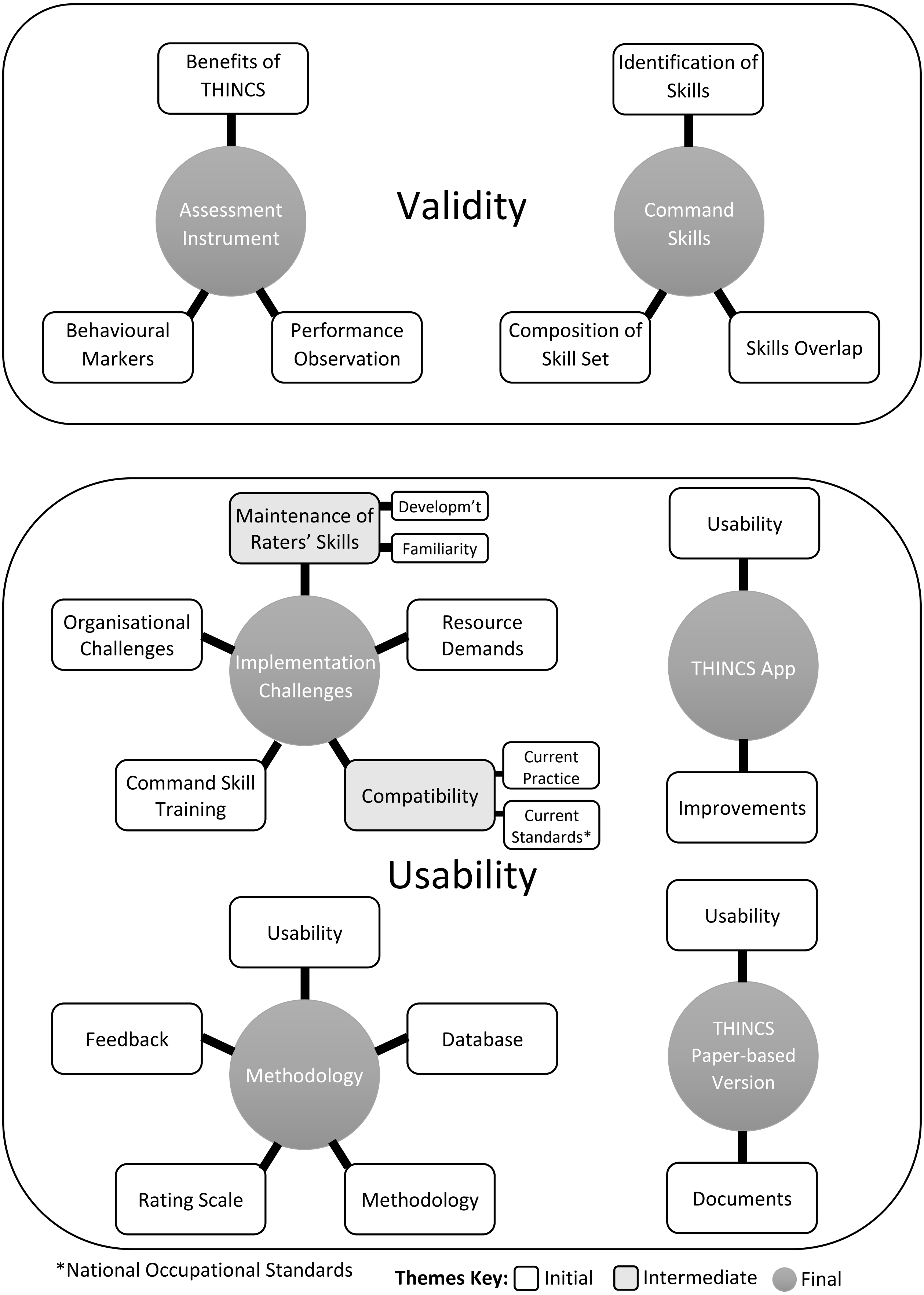
AOB

19. What is the worst/best thing about THINCS?

20. What is your worst/best experience of using THINCS?

21. Is there anything else you would like to say about THINCS?

Appendix 5: Focus Group Thematic Analysis Schematic



*National Occupational Standards

Appendix 6: Control Scenario Semi-structured interview questions

High Rise Semi-structured Interview Questions

Decision Point: Taking over/not taking over command

1. **What was the rationale for your decision?**
2. **When you took over/did not take over command what did you understand about the incident?**
    a. The situation?
        i. Have you had any command experience of this type of incident before?
    b. The resources?
        i. What was your rationale for increasing the resources (or did not)?
    c. The hazards and risks?
        i. What did perceive as the greatest hazards?
    d. What were you uncertain about?
        i. What were your information gaps?
        ii. Did you appreciate who was available and their skills and knowledge?
        iii. Did you fully trust the people involved?
        iv. Did you appreciate the capabilities of available appliances and equipment?
        v. Were you aware of how much time had passed at this point?

Decision Point: Initial Plan

3. **What was your plan at this stage?**
    a. How did you determine your objectives?
    b. How did you determine your priorities?
    c. What sources of knowledge were you relying on to determine your plan
    d. Were you relying on any SOPs, and if so, which ones?

Decision Point: Use of Initial Incident Commander

4. **How did you utilise the initial incident commander?**
    a. What was the benefit of using them in this way (or not using them)?
    b. What did you hope to achieve by using them this way?
    c. Have you used them in this way before?

Decision Point: Need for additional resources

5. **Why did you make up/Did you consider making up?**

a. How did you/would you intend to use the resources for?

b. What influenced/would have influenced the number and type of appliances?

c. What cues did/would you use to decide the scale of the make up?

Decision Point: Need to provide incident update

6. **Why did you send an informative message at this time/Did you consider sending an informative message?**

   a. What cues did/would you use to know when to send an informative?

   b. What sources of knowledge did/would you use to determine the message content?

   c. What rules did/would you follow?

Decision Point: Use of established bridgehead

7. **Why did you not commit/commit the BA crew from the bridgehead?**

   a. What was your rationale for the decision

   b. What cues did you use?

   c. What sources of knowledge did you use to help make the decision?

   d. What were you uncertain about?

   e. What SOPs/rules did you follow?

Decision Point: Use of the neighbour

8. **How did you use the neighbour?**

   a. What was the rationale for your decision?

   a. How did you come to that decision?

   b. What were you uncertain about?

   c. What was the benefit of your actions in relation to them?

   d. What did you hope to achieve by those actions?

   e. Have you done this before and were the circumstances similar?

Decision Point: Use of established bridgehead

9. **Why did you not commit/commit the BA crew from the bridgehead?**

   a. What was your rationale for the decision?

   b. What were you uncertain about?

   c. What cues did you use?

   d. What sources of knowledge did you use to help make the decision?

   e. What SOPs/rules did you follow?

Decision Point: Use of the HART resources

10. **How did you utilise HART?**
    a. What was the rationale for your decision?
    b. What were you uncertain about?
    c. What was the benefit of using them in this way (or not using them)?
    d. What did you hope to achieve by using them this way?
    e. Have you used them in this way before?
    f. What sources of knowledge did/would you use?
    g. What were the rules/SOPs you were following?

Decision Point: Use of the ALP

11. **How did you utilise the ALP?**
    a. What was the rationale for your decision?
    b. What were you uncertain about?
    c. What was the benefit of using it in this way (or not using it)?
    d. What did you hope to achieve by using it this way?
    e. Have you used it in this way before?
    f. What sources of knowledge did/would you use?
    g. What were the rules/SOPs you were following?

Decision Point: Use of the police resources

12. **How did you utilise the Police?**
    a. What was the rationale for your decision?
    b. What were you uncertain about?
    c. What was the benefit of using them in this way (or not using them)?
    d. What did you hope to achieve by using them this way?
    e. Have you used them in this way before?
    f. What sources of knowledge did/would you use?
    g. What were the rules/SOPs you were following?

Decision Point: Use of established bridgehead

13. **Why did you not commit/commit the BA crew from the bridgehead?**
    a. What was your rationale for the decision?
    b. What were you uncertain about?
    c. What cues did you use?
    d. What sources of knowledge did you use to help make the decision?

e. What SOPs/rules did you follow?

Decision Point: Use of upset relative of disabled resident

14. Why did you handle the upset relative in that way?

a. What was the rationale for your decision?
a. What were you uncertain about?
b. What was the benefit of your actions in relation to them?
c. What did you hope to achieve by those actions?
d. Have you done this before and were the circumstances similar?
e. Did you consider exploiting their relationship when making this decision, e.g. to update her relative?

Decision Point: Expansion and/or adaption of initial plan

15. When the extra pumps arrived how did that change your plan?

a. How do you feel the incident has developed?
b. How did you determine your objectives?
c. How did you determine your priorities?
d. What were you uncertain about?
    i. What were your information gaps?
    ii. Did you appreciate who was available and their skills and knowledge?
    iii. Did you fully trust the people involved?
    iv. Did you appreciate the capabilities of available appliances and equipment?
    v. Were you aware of how much time had passed at this point?
e. What information has influenced the adaptation of your plan?
f. Were there any specific cues?
g. How do you feel the risks have changed?
h. What sources of knowledge were you relying on to do this?
i. Were you relying on any SOPs, and if so, which ones?
j. What were you uncertain about?

Decision Point: Use of the ALP

16. How did you utilise the ALP?

a. What was the rationale for your decision?
b. What were you uncertain about?
c. What was the benefit of using it in this way (or not using it)?

d. What did you hope to achieve by using it this way?
e. Have you used it in this way before?
f. What sources of knowledge did/would you use?
g. What were the rules/SOPs you were following?

Decision Point: Use of established bridgehead

17. Why did you not commit/commit the BA crew from the bridgehead?

a. What was your rationale for the decision
b. What were you uncertain about?
c. What cues did you use?
d. What sources of knowledge did you use to help make the decision?
e. What SOPs/rules did you follow?

Decision Points: Response to the stress question:

1. How were you feeling at this point?
2. What were you uncertain about?
3. What cues influenced your level of stress at this time?
4. What made it increase/decrease from before?
5. How comfortable at this point are you with your decisions?
6. How were you managing your stress?
7. Were you relying on your training?
8. Was that level of stress affecting your ability to command, and if so, how?
9. Did that level of stress affect your behaviour, or other aspects of your performance?

Decision Points: Applies (or omits to apply) operational discretion appropriately or inappropriately

1. Were you aware of going outside of standard operational procedures?
2. What were you uncertain about?
3. What influenced your decision to do that/not do that?
4. What cues did you use?
5. What sources of knowledge were you relying on?
6. What is the procedure you should have followed?
7. Why did you stick to using SOPs?
8. Does your FRS have an Operational Discretion SOP?
9. Why did you not follow it?
10. Were there any organisational or cultural matters that influenced your approach?

Appendix 7: Discretion Scenario Semi-structured interview questions

Sinkhole Simulation Semi-structured Interview Questions

Decision Point: Taking over/not taking over command

1. **What was the rationale for your decision?**
2. **When you took over/did not take over command what did you understand about the incident?**
    a. The situation?
        i. Have you had any command experience of this type of incident before?
    b. The resources?
        i. What was your rationale for increasing the resources (or did not)?
    c. The hazards and risks?
        i. What did perceive as the greatest hazards?
    d. What were your information gaps?
        i. Did you appreciate who was available and their skills and knowledge?
        ii. Did you fully trust the people involved?
        iii. Did you appreciate the capabilities of available appliances and equipment?
        iv. Were you aware of how much time had passed at this point?

Decision Point: Initial Plan

3. **What was your plan at this stage?**
    a. How did you determine your objectives?
    b. How did you determine your priorities?
    c. What sources of knowledge were you relying on to determine your plan
    d. Were you relying on any SOPs, and if so, which ones?

Decision Point: Use of Initial Incident Commander

4. **How did you utilise the initial incident commander?**
    a. What was the benefit of using them in this way (or not using them)?
    b. What did you hope to achieve by using them this way?
    c. Have you used them in this way before?

Decision Point: Need for additional resources

5. **Why did you make up/Did you consider making up?**
    a. How did you/would you intend to use the resources for?

b. What influenced/would have influenced the number and type of appliances?

c. What cues did/would you use to decide the scale of the make up?

Decision Point: Need to provide incident update

6. **Why did you send an informative message at this time/Did you consider sending an informative message?**
    a. What cues did/would you use to know when to send an informative?
    b. What sources of knowledge did/would you use to determine the message content?
    c. What rules did/would you follow?

Decision Point: Use of the parents

7. **How did you handle and use the parents?**
    a. What was the rationale for your decision?
    b. What were you uncertain about?
    f. How did you come to that decision?
    g. What was the benefit of your actions in relation to them?
    h. What did you hope to achieve by those actions?
    i. Have you done this with parents before and were the circumstances similar?

Decision Point: Use of the line rescue tactical adviser

8. **How did you utilise the tactical adviser?**
    a. What was the rationale for your decision?
    b. What were you uncertain about?
    c. What was the benefit of using them in this way (or not using them)?
    d. What did you hope to achieve by using them this way?
    e. Have you used them in this way before?
    f. What sources of knowledge did/would you use?
    g. What were the rules/SOPs you were following?

Decision Point: Use of the police resources

9. **How did you utilise the Police?**
    a. What was the rationale for your decision?
    b. What were you uncertain about?
    c. What was the benefit of using them in this way (or not using them)?
    d. What did you hope to achieve by using them this way?
    e. Have you used them in this way before?

f. What sources of knowledge did/would you use?

g. What were the rules/SOPs you were following?

Decision Point: Use of the HART resources

10. How did you utilise HART?

a. What was the rationale for your decision?

b. What were you uncertain about?

c. What was the benefit of using them in this way (or not using them)?

d. What did you hope to achieve by using them this way?

e. Have you used them in this way before?

f. What sources of knowledge did/would you use?

g. What were the rules/SOPs you were following?

Decision Point: Expansion and/or adaption of initial plan

11. What was your plan at this stage?

a. How do you feel the incident has developed?

b. How did you determine your objectives?

c. How did you determine your priorities?

d. What information has influenced the adaptation of your plan?

e. What were you uncertain about?

 i. Did you appreciate who was available and their skills and knowledge?

 ii. Did you fully trust the people involved?

 iii. Did you appreciate the capabilities of available appliances and equipment?

 iv. Were you aware of how much time had passed at this point?

f. Were there any specific cues?

g. How do you feel the risks have changed?

h. What sources of knowledge were you relying on to do this?

i. Were you relying on any SOPs, and if so, which ones?

Decision Point: Use of tree surgeon parent

12. Why did you handle the tree surgeon parent in that way?

f. What was the rationale for your decision?

a. What were you uncertain about?

g. What was the benefit of your actions in relation to them?

h. What did you hope to achieve by those actions?

i. Have you done this before and were the circumstances similar?

j. Did you consider their expertise when making this decision?

Decision Point: Response to the stress question:

10. How were you feeling at this point?
11. What were you uncertain about?
12. What cues influenced your level of stress at this time?
13. What made it increase/decrease from before?
14. How comfortable at this point are you with your decisions?
15. How were you managing your stress?
16. Were you relying on your training?
17. Was that level of stress affecting your ability to command, and if so, how?
18. Did that level of stress affect your behaviour, or other aspects of your performance?

Decision Point: Applies (or omits to apply) operational discretion appropriately or inappropriately

11. Were you aware of going outside of standard operational procedures?
12. What were you uncertain about?
13. What influenced your decision to do that/not do that?
14. What cues did you use?
15. What sources of knowledge were you relying on?
16. What is the procedure you should have followed?
17. Why did you stick to using SOPs?
18. Does your FRS have an Operational Discretion SOP?
19. Why did you not follow it?
20. Where there any organisational or cultural matters that influenced your approach?