Normal Operations Monitoring - A New Approach to Measuring and Monitoring Human and Safety Performance – Tested in Aviation Ground Operations

Author:
Raggett, Louise

Publication Date:
2017

DOI:
https://doi.org/10.26190/unsworks/3202

Normal Operations Monitoring

A New Approach to Measuring and Managing Human and Safety Performance Tested in Aviation Ground Handling

Louise Raggett

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

School of Aviation

University of New South Wales

September 2016

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: Raggett

First name: Louise

Other name/s: Claire

Abbreviation for degree as given in the University calendar: AVIAAR1900 – Aviation, Doctor of Philosophy (PhD)

School: Aviation

Faculty: Science

Title: Normal Operations Monitoring – A new approach to measuring and monitoring human and safety performance - tested in aviation ground operations

Abstract 350 words maximum: (PLEASE TYPE)

Unlike flight operations, which is now regarded as an 'Ultra-safe system' (Amalberti 2001, p. 109) aviation ground safety has languished behind the rest of the industry (Verschoor and Young 2011), with activities on the ramp now accounting for more than a quarter of all incidents (Balk and Bossenbroek 2010). In recent years, both 'damage to aircraft and harm to ground personnel have escalated' (Passenier 2015 p. 38) costing air carriers more than USD10 billion annually (GAO 2007)

Models and measures influential to aviation safety are reviewed for their possible contribution to the ground safety problem. The Threat and Error Management Model (TEM) and Line Operation Safety Audit (LOSA) method (Klinect, Murray, Merritt and Helmreich 2003) were identified as potential candidates and critically examined for their suitability. The aim of the current research was to develop an approach that builds on the existing benefits of LOSA, whilst incorporating contemporary theories of safety science, and improving data reliability and validity where possible.

A new method is proposed named 'Normal Operations Monitoring' (NOM). NOM is an observation methodology which codes and measures human and safety performance in routine operations. NOM attempts to measure the gap between the system as designed and the system as actually operated (Hollnagel 2007). Identifying this variance provides novel insights into factors which influence safety performance and suggest new opportunities for interventions and improvement.

NOM was customised for application in the ground handling industry. Data was collected and analysed from over 1300 observations of aircraft turnarounds. Implications for ground safety are explored as well as the potential applications and benefits of NOM to other domains. The final discussion explores how the current research and NOM tools could be taken forward as a method for informing and improving safety management in other high-hazard industries.

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed .

Date ..

COPYRIGHT STATEMENT

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.
I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).
I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed

Date

AUTHENTICITY STATEMENT

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed .

Date .

THE UNIVERSITY OF NEW SOUTH WALES

Thesis/Dissertation Sheet

Surname or Family name: Raggett

First name: Louise

Other name/s: Claire

AVIAAR1900 – Aviation

Doctor of Philosophy (PHD)

School: Aviation

Faculty: Science

Title: Normal Operations Monitoring - A New Approach to Measuring and Monitoring Human and Safety Performance – Tested in Aviation Ground Operations

TABLE OF CONTENTS

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Ann Williamson, for her continuous support of my PhD study, her patience, her motivation and her immense knowledge. In our time together Ann has taught me the meaning of scholarship. I could not have imagined having a better supervisor and mentor for my PhD. Besides my advisor, I would like to thank Professor Jason Middleton for his encouragement and support throughout this process and to the School of Aviation at UNSW for their support of this thesis.

This study was only possible because of the many people in the aviation industry who gave of their time to make it happen. From the leaders and managers who had the vision to allow the study to go ahead, to the enthusiastic team members who contributed their skill and enthusiasm, to the observers who had the patience to conduct hours of observations, I would like to express my sincere respect and gratitude. For reasons of confidentiality, I cannot name the individuals who were most instrumental in getting this work off the ground, but I hope they know who they are and how thankful I am for their time and effort.

I would also like to thank the late Professor John Wilson for encouraging me to start the PhD. in the first place, and believing I could finish it in two years. I am sorry that in the end I was not able to complete it in his lifetime. I will be forever grateful for his confidence in me, and for giving me a start. Unfortunately, it took much longer than anyone anticipated, and I must humbly thank my family for their seemingly endless patience.

Special thanks must go to my sister Lyndsey Regan and father Brian Raggett, who offered proofreading assistance and feedback when it was most needed. I really appreciate your help. To both my parents, I would like to thank you for a lifetime of academic support and encouragement to shoot for the stars. Last but not least, I must thank my husband, Mark Jackson, without whom none of this would have been possible. Your words of encouragement kept me going. You have been my rock since the beginning and my biggest fan at the end. Thank you for your support in all its forms, emotional, financial and practical.

Finally, to my children Emelia and Oscar Jackson, who were both born in the eight years it took to produce this piece of work, I apologise for the sacrifices made in pursuit of this goal. I hope, however, that if nothing else, I have modelled the life skills of persistence and perseverance that you might one day value. I promise to support your own goals in life with the same commitment and energy I have devoted to my own, whatever they may be.

List of Figures

Glossary

Aircraft turnaround: the process of unloading and reloading an aircraft whilst it is on the ground between flights or sectors.

Behavioural marker: A single non-technical skill or competency within a work environment that contributes to effective or ineffective performance.

Crew resource management (CRM): A team training and operational philosophy with the objective of ensuring the effective use of all available resources to achieve safe and efficient flight operations. (Sometimes described as Non- Technical Skills training or NTS)

Ground Servicing Equipment (GSE) - equipment used on the airport ramp to service the aircraft and loading of aircraft during the aircraft turnaround.

Human Error Identification (HEI): techniques for identifying the types of error that may occur often used as a part of Human Reliability Assessment (HRA)

Human Factors (HF): The minimisation of human error and its consequences by optimising the relationships within systems between people, activities and equipment

Human Reliability Assessment (HRA): techniques for evaluating the probability of a human error occurring throughout the completion of a specific task.

Inter-rater reliability: The extent to which two or more individuals (coders or raters) agree. Inter-rater reliability addresses the consistency of the implementation of a rating system.

Line Operational Safety Audit (LOSA): A LOSA is a formal process where highly trained observers ride the jump seat during regularly scheduled flights to collect safety-related data on environmental conditions, operational complexity, and flight crew performance. Confidential data collection and non-jeopardy assurance for pilots are fundamental to the process.

Non-technical skills: the mental, social, and personal-management abilities that complement the technical skills of workers and contribute to safe and effective performance in complex work systems. They include competencies such as decision-making, workload management, team communication, situation awareness, and stress management

Normal Operations Model- a new approach for monitoring and measuring system performance during normal routine operations

NOTECHS: A framework, designed under the Joint Aviation Authorities (Europe), for the structured assessment of non-technical skills, based on a behavioural marker system.

Ramp - A defined area on an airport intended to accommodate aircraft for loading or unloading of passengers, mail or cargo, or for fuelling, parking or maintenance [IATA ITRM, 2009].

Threat and error Management Model – The Threat and Error Management (TEM) framework is a conceptual model that assists in understanding, from an operational perspective, the inter-relationship between safety and human performance in dynamic and challenging operational contexts. (ICAO 2002)

Threat – In NOM a threat is defined situation or condition that occurs outside the influence of those being observed (i.e. it was not caused by the crew being observed), increases the operational complexity, and may require attention. It has the potential to increase risk to the operation.

Training Needs Analysis: The identification of training needs at employee, departmental, or organisational level, in order for the organisation to perform effectively.

Undesirable operational state (UOS) – An outcome that has the potential for consequential (unsafe outcomes) and is therefore defined as undesirable by the organisation

Undesired Aircraft State: An outcome in which the aircraft is unnecessarily placed in a compromising situation that poses an increased risk to safety (ICAO 2002)

Unit Load Device (ULD) - a pallet or container used to load luggage and freight on an aircraft, allowing a large quantity of cargo to be handled as a single unit.

Variation – in NOM a variation is an action or inaction by the team under observation, which varies from the work as described. An objective standard is needed to assess whether a variation has occurred by referring to standard such as a standard operating procedures or operations manuals

List of Acronyms

ADAMS	Aircraft Dispatch and Maintenance Safety
ASR	Air Safety Reports
ATSB	Australian Transport Safety Bureau
CAA NL	Civil Aviation Authority of the Netherlands
CAA UK	Civil Aviation Authority of the United Kingdom
CAAP	Civil Aviation Advisory Publication
CAO	Civil Aviation Order
CASA	Civil Aviation Safety Authority
CASR	Civil Aviation Safety Regulation
CAST	Commercial Aviation Safety Team
EASA	European Aviation Safety Agency
ECAST	European Commercial Aviation Safety Team
FAA	Federal Aviation Administration (United States)
FMAQ	Flight Management Attitudes Questionnaire
FOD	Foreign Object Debris
FSF	Flight Safety Foundation
GAO	United States Government Accountability Office
GHO/P	Ground Handling Organisation/Provider
GSE	Ground Service Equipment
GSP	Ground Service Provider(s)
HSE	United Kingdom Health and Safety Executive
IATA	International Air Transport Association
ICAO	International Civil Aviation Organisation
ISAGO	IATA Safety Audit for Ground Operations
LOSA	Line Operations Safety Audit
MEDA	Maintenance Error Decision Aid
NASA	National Aeronautics and Space Administration
NOM	Normal Operations Monitoring
NOSS	Normal Operations Safety Survey
NOTECHS	Non-Technical Skills
NTSB	National Transportation Safety Board
OTD	On Time Departure
REDA	Ramp Error Decision Aid SAT Safety Analysis Team
R-LOSA	Ramp Line Operations Safety Audit
SMS	Safety Management System
UT	University of Texas

Abstract

Unlike flight operations, which is now regarded as an 'ultra-safe system' (Amalberti 2001, p. 109), aviation ground safety has lagged behind the rest of the industry (Verschoor and Young 2011), with activities on the ramp now accounting for more than a quarter of all incidents (Balk and Bossenbroek 2010). In recent years, both 'damage to aircraft and harm to ground personnel have escalated' (Passenier 2015, p. 38), costing air carriers more than USD10 billion annually (GAO 2007)

Models and measures influential to aviation safety are reviewed for their possible contribution to the ground safety problem. The Threat and Error Management (TEM) model and Line Operation Safety Audit (LOSA) method (Klinect, Murray, Merritt and Helmreich 2003) were identified as potential candidates and critically examined for their suitability. The aim of the current research was to develop an approach that builds on the existing benefits of LOSA, whilst incorporating contemporary theories of safety science and improving data reliability and validity where possible.

A new method is proposed, named 'Normal Operations Monitoring' (NOM). NOM is an observation methodology that codes and measures human and safety performance in routine operations. NOM attempts to measure the gap between the system as designed and the system as actually operated (Hollnagel 2007). Identifying this variance provides novel insights into factors that influence safety performance and suggest new opportunities for interventions and improvement.

NOM was customised for application in the ground handling industry. Data was collected and analysed from over 1300 observations of aircraft turnarounds. Implications for ground safety are explored as well as the potential applications and benefits of NOM to other domains. The final discussion explores how the current research and NOM tools could be

taken forward as a method for informing and improving safety management in other high-hazard industries.

Chapter 1: Overview of Thesis

This research begins with the assertion that there are currently no suitable methods for collecting data about human and safety performance in aviation ground operations to drive safety improvement. It is argued that this lack of data is currently contributing to the stagnation of safety improvement in this sector. The aim of this research is to develop and test a new approach to collecting and analysing data about human and safety performance in aviation ground operations. Managers recognise it is difficult to manage safety without measuring it; however, knowing what to measure is a question that has perplexed safety science for some time. Since the dawn of the industrial age, theorists have searched for explanations about how accidents occur in the hopes that, by understanding accident causation, the next accident can be prevented. Models of accident causation have inspired the measurement of many different indicators of safety, but none have been universally accepted. Modern risk management practices recognise that accidents arise out of the complex interactions that occur between people, organisations and the hazards they manage. A solution for how to measure and manage safety must, therefore, take account of the interplay between humans and the operational system.

Typically, organisations have many safety indicators but may have difficulty identifying which measures are important for reflecting the current safety status, or changes to that status that require attention to keep the system safe. While it is now widely recognised that human performance is crucial to safety outcomes, surprisingly few organisations routinely measure human performance in their operations. In order to improve safety, more information is needed about how people manage hazards and risks and, equally, how the system itself impacts on human performance.

Measures are needed that provide data congruent with modern safety management practices such as Safety Management Systems. Today's safety management systems are based on the assumption that system performance data provides feedback loops, so systems can continuously adjust and improve their performance. If systems are unable to adjust and improve, it may be because the data is not providing the right feedback to drive continuous improvement, indicating that new measures are now needed.

One industry where safety performance improvement appears to have stalled is aviation ground safety. Unlike the rest of aviation, which has continuously improved to what Amalberti refers to as an '[u]ltra-safe system' (Amalberti 2001, p. 109), ground safety has languished behind the rest of the industry (Verschoor and Young 2011). Activities on the ramp now account for more than a quarter of aviation incidents (Balk and Bossenbroek 2010). In recent years, both 'damage to aircraft on the ground as well as harm to ground personnel have escalated' (Passenier 2015, p. 38).

Ground operations play an important role in the aviation industry, yet have received little attention in safety or human factors research. Risks created here have caused aircraft depressurisation, structural damage, corrosion and fire. Incorrect loading has also lead to aircraft handling accidents, tail strikes, rejected take-offs and injury to personnel (Marchbank 2009, Boer et al. 2012). Despite the risks and costs involved, ground operations have not been the subject of much academic study. Research into aviation safety and human factors has, to date, focused largely on the safety of flight. The aim of the current research is therefore to review current approaches to human and safety measurement and develop a method suitable for application in the ground safety environment. The aim will be to identify indicators that inform existing safety management systems, with a view to developing evidenced-based interventions that will reinvigorate the continuous improvement of ground safety. It is hoped that, with better data, understanding and feedback, ground safety may one day reach parity with the safety record the rest of the industry experiences.

The research begins with an exploration of the ground safety problem and the issues the industry is facing today. This is followed by a review of models of accident causation that have shaped our understanding of how safety is created and maintained, and influences the data collected about safety performance. This is followed by a review of the measures available for measuring safety and a discussion of their relative advantages and disadvantages. Elements from existing models and methods are adapted to propose a new approach for measuring and monitoring safety, using an observation methodology.

A framework is proposed to establish the criteria that a new data collection tool should meet and the standard by which it will be evaluated upon completion. The rest of the thesis then charts the development of the method and proposes a new approach named Normal Operations Monitoring, or NOM. NOM is then customised for application in ground handling, as a test of the method and its ability to provide data to inform new strategies for improving safety. Finally, NOM is assessed against the proposed evaluation framework to determine its usefulness for collecting new and unique data with which to drive continuous improvement.

The results show that, although developed and tested in a ground handling environment, NOM is a generic tool that could be applied anywhere new information is needed about human and safety performance. The final discussion explores ways in which the current research and NOM tools could be taken forward as a method for informing and improving safety management in high-hazard industries generally.

An overview of the structure and content of the thesis is provided in the Chapter summary below.

Chapter 2 begins by exploring the impact of ground incidents on the aviation industry. Evidence is explored that demonstrates the impact of ground incidents in terms of cost, aircraft damage, injury to personnel and safety of flight incidents (Lacagnina 2007). In order to move the ground handling industry forward from its current safety performance, it is argued that new methods of collecting and analysing data in this field are urgently needed.

Chapter 3 explores models of accident causation that have helped to determine what information is important to measure and improve safety performance. The models for organising and collecting data are explored. The benefits and limitations of past and present safety concepts and models are examined, as well as their influence on the aviation industry to date, with a view to identifying concepts helpful for the current research. Several potential models are presented and discussed. Elements from theoretical models are identified as beneficial to take forward for the current study, including concepts from the Threat and Error Management (TEM) model of human performance (Klinect, Wilhelm and Helmreich 1999) and the Reason model of accident causation in organisational systems (Reason 1995).

Chapter 4 explores how the models reviewed in Chapter 3 have shaped what data is collected to monitor and measure safety. Methods for data collection and measurement are reviewed. Benefits and limitations of the approaches are discussed, as well as their application in the aviation industry to date. The objective is to identify desirable features of a new data collection tool for ground safety. It is proposed that observation may provide the best opportunity to collect data about naturalistic behaviour, in real-world situations, from the perspective of the operator. The Line Operation Safety Audit, or LOSA, which is based on the TEM model explored in Chapter 3 (Klinect, Murray, Merritt and Helmreich 2003) is explored as an observation tool used in aviation that may be suitable to take forward in the current research.

Chapter 5 offers a deeper review of the LOSA methodology and TEM as its underlying conceptual model. This chapter provides a review of the origins of the LOSA methodology and the development of its component parts. The proliferation of the method, both within aviation and beyond, is reviewed along with its proposed benefits, including a recent adaptation for use in ground operations. Despite its popularity and apparent suitability, however, serious concerns are raised regarding the validity of LOSA and whether it measures what it purports to measure. Theoretical questions and concerns regarding validity are explored. It is concluded that while many elements of the LOSA method could be useful to this study, there are a number of characteristics of the model and method that would require development to address the methodological concerns. Further investigation is recommended to identify which aspects of the tool should be retained or redesigned for the purposes of the current research.

Chapter 6 proposes an evaluation framework for the assessment of LOSA's strengths and weaknesses as a data collection tool. The benefits of evaluation frameworks are briefly discussed, as are the criteria against which a tool or methodology can be assessed. In the area of human performance, one of the most comprehensive evaluation frameworks was developed by Kirwan (1992, 1996 and 1998). Barber and Stanton (1996) also provide an example of how Kirwan's original principles can be adapted for evaluating methodological tools. In the current research, both Kirwan's (1992b) and Barber and Stanton's (1996) criteria are used as the basis of a bespoke framework for evaluating LOSA. The adapted criteria are then discussed in relation to LOSA as it is today, to identify features of LOSA that should be retained and those that should be modified. The aim is to establish the criteria upon which to design the new data collection tools for ground operations, and eventually assess the resulting tool against the same evaluation criteria to judge its effectiveness.

Chapter 7 describes how the development needs identified in Chapter 6 are incorporated into the adaptation of a new method based on the LOSA principles. This chapter proposes a new model of human performance based on TEM and LOSA. In the new model, changes to terminology are proposed. 'Threats' to safety are coded in a similar way to the LOSA method but, rather than classifying errors, it is proposed that 'variations' to standard operating procedures are recorded instead. Safety outcomes are noted, including the occurrence of 'undesired operational states' (UOSs). The observation method itself is restricted to noting observable actions and outcomes. Relationships between elements in the model are inferred not through observation but through post-hoc statistical analysis. The new method, known as Normal Operations Monitoring (NOM), is developed to be generic in nature, suitable for application in any industry where data is sought about human and safety performance to drive continuous improvement.

Chapter 8 then describes the application of NOM within live ground handling operations. The adaptation of NOM codes and tools for the ground handling domain is outlined. The development and refinement of the method is described. Loading supervisors were trained in the new NOM observation model, method and coding system and asked to observe aircraft turnarounds. The study recorded over 1300 observations in 28 ports internationally. The data verification and analysis process is described and the results presented in chapters 9 and 10.

Chapter 9 provides a general overview of the data. Data is presented with regard to the most frequently occurring threats, variations from procedures and undesired operational states. Each of the major categories of threats and variations is explored in more detail to understand their contribution. The number and type of undesired states are described. Finally, Chapter 9 reviews how well each threat, variation and undesired operational state was managed by the ramp crew and compares this data to LOSA data in other domains, where data is available.

Chapter 10 looks in more detail at the relationships between elements of the NOM model. This was achieved by first looking for significant associations between categories of threats and variations as well as associations between variations and undesired operational states. Significant associations are then explored in more detail. The data is able to identify, for example, which threats are associated with which variations and with which undesired operational states, through a series of diagrams. The implications for ground safety in terms of the high-value targets for safety interventions and improving organisational defences are discussed. An amended NOM model is presented based on lessons learned from the study. The implications for ramp safety and for the NOM model are discussed.

Chapter 11 assesses the new NOM tool against the evaluation framework proposed in Chapter 6. The success of NOM is considered in more detail against the criteria proposed for its development. The evaluation of NOM is compared against its LOSA predecessor and found to offer favourable improvements in a number of areas. In summary, when considered in terms of the evaluation criteria developed in Chapter 6, NOM is considered to meet its development objectives of building on the strengths of LOSA and making modifications to overcome its shortcomings.

Chapter 12 discusses the results of the study overall, including whether the research has met what it set out to achieve, namely to provide a new method for collecting data about human and safety performance in ground operations. It is argued that NOM data can provide many benefits, such as informing evidence-based safety interventions and providing an ongoing measure of their effectiveness. The data collected is novel and insightful in terms of providing high-value targets for improving safety. Implications for ground handling are discussed, as well as the possible applications for NOM in other domains. Potential benefits of NOM are outlined, as well as avenues for future research. Chapter 12 concludes that NOM provides an effective new tool for measuring a human and safety performance in ground operations or indeed any high-hazard environment.

Chapter 2: Safety and Human Performance in Ground Operations

Ground handling is an important part of the aviation sector that is often overlooked. The global airport ground handling business is estimated to be worth over USD80 billion per annum according to its trade association, yhe Airport Services Association (ASA) (Ornellas 2015) with more than 1000 ground handling companies operating worldwide.

According to the International Air Transport Association (IATA), ground safety events occur so frequently and result in such expensive damage that ramp accidents are often the second highest cost to airlines after fuel costs (Negroni 2001). It is surprising that although incidents on the ground cost the aviation industry billions of dollars a year, ground safety has received relatively little attention from the academic community. This chapter will outline the impact human performance issues are having on the aviation industry today, and argue why more research in this area is urgently needed.

2.1 The Ground Safety Problem

The airport ramp can be a dangerous place. The Australian Transport Safety Bureau (ATSB) has recognised this, saying: 'Ground operations are potentially one of the most dangerous areas of aircraft operation' (2010). Indeed, more than one quarter of all aircraft incidents occur on the ground (Boer, Koncak, Habekotte and Van Hilten 2012). Ground incidents contribute to aviation accidents, human injuries and financial strains.

Since the 1990s, the aviation industry been experiencing an overall decrease in incident rates (Ananda, Kumar and Ghoshhajra 2010, Flouris and Kucukyilmaz 2009, Johnson, Kirwan, Lieu, and Stastny 2009, Layton 2012, Mitchell and Leonhardt 2010, Duncan 2013). However, despite a consistent overall reduction of aviation safety incidents per passenger mile, the number of ground occurrences has remained unimproved since 2000 (ATSB 2010, p. 4). The International Civil Aviation Organisation's (ICAO) reports have shown that the rate of ground accidents is, in fact, increasing (ICAO 2013, p. 36). Based on data compiled by the International Air Transport Association (IATA), the FSF estimates that 27,000 ramp accidents and incidents – one per 1000 departures – occur worldwide every year, injuring about 243,000 people a year (FSF 2007).

In its 2014 'Global Safety Aviation Study,' the Allianz Global Corporate and Specialty group identified a number of factors that might be contributing to the issue: 'Ramp congestion, increasing numbers of flights, stringent aircraft scheduling requirements and efforts to position large jets into gates originally designed for much smaller aircraft (contribute to) traffic jams and tight quarters on the ramp' (p. 41). The situation is likely to get worse as airports become even more congested.

Both Airbus (2013) and the Federal Aviation Administration (FAA 2012) have suggested that air traffic will double in the next fifteen years, planes will increase in size, airports will get busier and schedules will be even tighter. Given these pressures, the next frontier in aviation safety is not likely to present itself in the air but on the ground and, as Matthews predicted (2012), without reform 'we will not see the improvement in safety that the rest (of the aviation) industry has come to expect' (p. 2).

The Flight Safety Foundation (FSF)'s Ground Accident Prevention (GAP) program was established in 2003 in response to requests from member organisations to find ways of improving ramp safety conditions. The GAP subsequently found that human factors – in particular, noncompliance with standard operating procedures – contributed largely to ramp accidents and incidents (as cited by Allianz 2014, p. 42). Human errors on the ramp reduced overall safety and increased the number of flight incidents associated with incorrectly loaded aircraft, aircraft damage and ramp crew injuries. The delays caused by such incidents, in and of themselves, can cost airlines thousands of dollars per minute.

Whilst research into aviation safety and human factors in particular has focused largely on the safety of flight, human factors in ground operations have not been the subject of much research. Drury (2000) reported that efforts made by airlines, consultants and human factors professionals had been focused, in the preceding 25-year period, almost exclusively on reducing errors on the flight deck, improving pilot training and, to a lesser extent, identifying maintenance errors. This suggests that significant research is now needed in order to meet the challenges of ground safety facing the industry today and in the future.

The remainder of this chapter explores in more detail the nature of ground operations and the safety challenges facing the industry today. It also explores potential reasons for the relative scarcity of research to date and explains how this knowledge gap can be addressed, suggesting new approaches in human factors research.

2.2 Ground Safety Incidents

2.2.1 Safety-of-flight impacts

A primary aim of ground handling tasks is to ensure the safe loading of aircraft. Aircraft loading errors can cause aircraft configuration and handling difficulties, which can affect an aircraft's centre of gravity, sometimes referred to as incorrect trim. Weight and balance issues can cause serious problems at take-off and for those attempting to control the craft while in flight. Since pilots calculate fuel requirements based on loaded aircraft weight, errors in loading can also result in incorrect fuel ratios. Similarly, incorrect load positioning in the hold, can contribute to aircraft damage or handling problems, as cargo moves and weights shift while in transit.

An ATSB (2010) study into Aircraft loading occurrences reports that serious incidents result from shifting cargo and unlisted cargo being loaded affecting aircraft controllability. In addition, incorrectly calculated weights have also been responsible for causing serious aviation incidents during take-off and landing (ATSB 2010, p. 1). Undetected problems can include ‘nose heavy’ outcomes, where excessive back pressure is required on the control column to lift the aircraft from the ground, or alternatively ‘tail heavy’ outcomes, which can result in the aircraft tail striking the ground (ATSB 2010, p.2). For example, in 2009 an Airbus A320 suffered a tail strike during take-off from Verona, Italy, because of an uneven load. Similarly, in 1999 in Guernsey in the UK, a tail-heavy F27 aircraft pitched up significantly as the wing flaps were applied on final approach, when the cargo shifted causing the aircraft to stall (ATSB 2010, p. 3).

Marchbank (2009) offered further examples of safety-of-flight incidents that can result from damage caused during aircraft loading, in the form of excerpts from the U.S. government’s National Transportation Safety Board (NTSB) accident database (2008). In one such example, an Alaska Airlines flight departing from a Washington state airport was said to have been ‘rocked by a “thunderous blast” and dropped from 26,000 ft’, as a result of a crease (or dent) in the side of the aircraft made by baggage handlers whilst manoeuvring loading equipment. The MD-80 aircraft returned to Seattle and, upon inspection, it was found that the original crease had developed into a 1.5–ft hole 20 minutes after take-off, which had led to a depressurisation of the aircraft.

In another example (Marchbank 2009), a Northwest Airlines DC-9-31 was climbing through 20,000 ft when its crew members heard what they later described as ‘a loud pop’ followed by depressurisation of the cabin. The flight landed without further incident, but the NTSB probable cause report notes that ‘a post-flight inspection revealed a 12-in.-by-5-in. fuselage skin tear, approximately six feet forward of the forward cargo door’, which resulted from a senior ground agent’s attempts to move the belt loader away from the aircraft by pushing it with a tug. The incident was initially not reported, as the agent purportedly advised co-workers not to say anything (NTSB 2008).

Under-reporting of ground safety incidents is also a serious concern. Ground damage events that may appear insignificant at the time can subsequently affect the safety of flight. This is of particular concern for the new, lighter, aircraft made of composite materials (Allianz Global Claims Review 2014) Newer generation aircraft, such as the Airbus A350 and the Boeing 787 Dreamliner, are built using composite carbon fibres, encased in toughened resins. Pierobon (2014) describes how composite structures can create an increased risk because of the behavioural properties of these materials when impacted. The surface of the material can return to its original shape after being hit, which may conceal damage underneath (Tomblin 2006). This can result in even apparently minor dings affecting the future structural integrity of the aircraft (Baker 2004). In September 2011, the U.S. Government Accountability Office (GAO) released a report highlighting the challenges in detecting damage in composite materials, and in predicting how such materials will behave when damaged. Balk's (2008) study found that in some cases where damage is found on an aircraft, no cause can be specified because the incident was not reported. Balk concluded that 'unreported damage' (p. 28) may pose the highest risk to flight safety, where risks to structural integrity of the aircraft are not routinely reported. With the increase of composite materials in the future, preventing and reporting ground damage may become even more imperative.

2.2.2 Injury to ramp personnel

While safety-of-flight incidents have the potential to contribute to the loss of multiple lives, loss of life and serious injury related to ground incidents are far more common on the ground than in the air. When analysing the type of injuries that occur on the ramp, Marchbank (2009) noted that most injuries occurring on the ramp stem from vehicle collisions, jet blast, manual handling incidents and fires that ignite during refuelling, along with other risks associated with moving vehicles and falls.

In November 2007, the U.S. GAO issued a report on ramp safety that showed that, in the preceding five years, fatal airport ramp accidents occurred (in the U.S.) at a rate of six per year. This led the International Air Transport Association (IATA) to conclude in 2009 that the airport ramp is one of the most dangerous workplaces in the world (Negroni 2011). Learmont similarly noted that, while 'commercial aviation may be justifiably proud of its safety in the air, its industrial injury record on the ground is one of the worst among all businesses' (2005, p. 1). In support of that statement, Lacagnina (2007) pointed to FSF estimates that suggest 243,000 airline industry personnel are injured on the ramp every year. Piotrowicz, Edkins and Pfister (2002) have also noted that workplace personal injury rates on the airport ramp for one Australian carrier were '20 times greater than injuries occurring in flight operations' (p. 2).

Those working within the FSF's GAP have gone so far as to say that injury rates related to ramp work closely resemble injury data associated with the coal mining and timber industries (Kerkloh 1992, Kerkloh and Tumeulty 2002). Indeed, ground handling occupational health and safety data measures up poorly when held against comparable workplaces in which manual tasks prevail. Vandal (2004), upon reviewing data from the U.S. Bureau of Labor, suggested that the ramp is a more dangerous place to work than a coal mine or a construction site, saying: 'Scheduled commercial air transport has 10.5 total recordable injuries per 100 employees – higher than all other industries, including oil and gas (2.5), pulp and paper manufacturing (3.2), the chemical industry (3.5) even construction (6.4)' (p. 17). Vandal noted that scheduled commercial air carrier performance was considerably worse in the case of both total recordable injuries and lost workdays – with the majority of these injuries occurring on the airport ramp.

In the U.K., an HSE report analysing the period 2009–2013 showed that injury trends within the U.K.'s airport transport industry had not improved. Similar data in Australia suggests the situation here is not much better. Data compiled by Safe Work Australia (2010) shows that ‘services to air transport, recorded the largest percentage increase in the number of serious claims (or 143%) over the period (of 2005 to 2010)’. In addition, increases in frequency rates were also recorded for services provided to air transport.

2.2.3 Damage to aircraft

A ground damage incident refers to any incident that results in damage while the aircraft is on the ground – such as a collision with other aircraft, ground vehicles or ground servicing equipment. Balk (2008) observed ground interactions at Amsterdam’s Schiphol airport and suggested, from his analysis, that ground damage due to human error is relatively common. While noting that ground damage occurred at ‘a rate of one ground handling incident with resulting aircraft damage per 5000 flights’, he estimated that 90 per cent of all ground damage incidents are caused by the actions of ground personnel engaged in ground handling activities. (p. 23).

2.2.4 Calculating the cost to industry

Injuries and aircraft damage caused on the ground add significantly to both direct and indirect industry costs each year. Direct costs include costs related to personnel injury as well as damage to aircraft, airport facilities and ground-support equipment. Indirect costs include replacement of personnel and parts, damage to reputation and increases in insurance policy premiums.

Several studies have contributed to today’s understanding of how such costs arise. Matthews (2012), for example, used U.S. injury compensation cost data to estimate initial health-related expenses and concluded that – with average costs of health related expenses ranging from USD173 to USD571 per worker, for a typical American airline – it is possible to project that costs could run into the ‘tens of millions per year per company’ (p. 4).

When estimating costs related to ramp damage to aircraft, Matthews used data from three representative airlines of differing sizes. After excluding uncontrollable incidents (i.e.

those caused by bird strikes, hail damage and lightning), Matthews's estimates for incidents involving human error ranged from USD77,000 to USD93,000 per aircraft. Matthews further hypothesised that, if the average cost per event totaled as little as USD50,000, and with a world fleet of 12,300 aircraft, *per carrier* direct costs related to ground incidents would total USD615 million annually.

The safety committee for the National Business Aviation Association (NBAA) estimated that ground mishaps cost the aviation industry more than USD100 million annually (2012). Other estimates suggest the impact is even greater. The IATA reported in its 2015 annual review that 'Ground damage costs airlines an estimated [USD]4 billion per year' (IATA Annual Review 2015, p. 32) when considering the associated indirect costs. If indirect costs are taken into account, the costs may be higher still.

The *Ground Support Worldwide* magazine reports that, when airlines factor in costs of the aircraft being out of service, increased insurance premiums, temporary replacement, employee injuries, and other associated expenses, the direct costs associated with aircraft damage on ground estimated costs can approach USD5 billion a year (Werfelman 2009). A U.S. GAO report put the number even higher, suggesting that that ground mishaps worldwide could be costing air carriers more than USD10 billion annually in aircraft damage, delays and associated indirect expenses (United States Government Accountability Office Report 2007). The FSF (2009) came up with a similar total figure. In its estimate, the FSF calculated annual worldwide costs to be approximately USD4.2 billion in ground incidents, aircraft damage and associated costs, and USD5.8 billion in injuries. In total, those sums equal nearly USD10 billion dollars in annual direct and indirect costs.

2.2.5 Identifying the causes of ramp incidents

Unfortunately, the ground safety industry has lacked standardised data to indicate what the problems are and where they are occurring. U.S. government reports have acknowledged problems with regard to ramp safety that are 'hindered by the lack of data on the nature, extent, and cost of such incidents and accidents' (GAO 2012, p. ii). It notes that successful safety strategies depend on having a thorough understanding of the problem and its causes. Unfortunately, the ground safety industry suffers from a lack of standardised data that would indicate what the problems are and where they are occurring.

Previous attempts have been made to standardise data collection (e.g., the Aerospace Psychology Research Group, IATA ground accident database and the SCARF program); however, Matthews (2012) notes that there are still many gaps in the data and therefore our understanding of how to make improvements (p. 2). Landry and Ingolia's (2011) study into ramp safety practices also concluded that more information is necessary to be able to understand the nature and extent of the problem as a first step to identifying what actions are needed to reduce ramp accidents and incidents.

2.3 Understanding the Ground Handling Task

The ground handling task takes place in a complex contractual, economic and regulatory environment, which may be contributing to its poor safety performance and confounding the industry's progress in achieving safety improvements. These issues will be explored in more detail below.

2.3.1 The ground servicing task

IATA defines ground handling as 'complex series of processes required to separate an aircraft from its load (passengers, baggage, cargo and mail) on arrival and combine it with its load and fuel prior to departure' (IATA, ISAGO Standards Manual 2016, p. 7) Ground activities take place in congested spaces, often in compressed timeframes and with multiple tasks occurring simultaneously (2007). Keeping an aircraft's time on the ground to a minimum is a critical factor for aviation profit margins. This means that loading must be performed at the same time as maintenance and servicing of the craft. The window of time in which the aircraft is on the ground between flights is commonly referred to as its turnaround. This window of time can range from 20 minutes to two hours – depending on the size and destination of the aircraft – but is typically less than 45 minutes even for very large aircraft.

The number of teams requiring access to the aircraft during this window is extensive and can include cargo/ baggage handlers, engine refuellers, potable-water refuellers, toilet systems servicers, catering personnel, cleaners, aircraft servicers, maintenance engineers, pushback drivers, load controllers, aircraft washers, de-icing teams, freight handlers, construction workers, aircraft maintenance staff and security personnel. At times, cabin crew and passengers are also moving in and around the aircraft during turnaround. In addition, there is traffic generated by ground service personnel attempting to access other aircraft.

The need to manage multiple tasks within the compressed time-window of a turnaround has been compared to the action in a Formula One pit-stop (Coleman 2009). However, unlike the highly choreographed pit-stop, each ramp servicing team may be employed by a different company, with different, sometimes conflicting priorities. Coordination between these interfaces can therefore be difficult, with a need for precision timing and coordinated activity. Human performance during this critical period, therefore, is vital to the success and safety of the operation overall. (Figure 1).

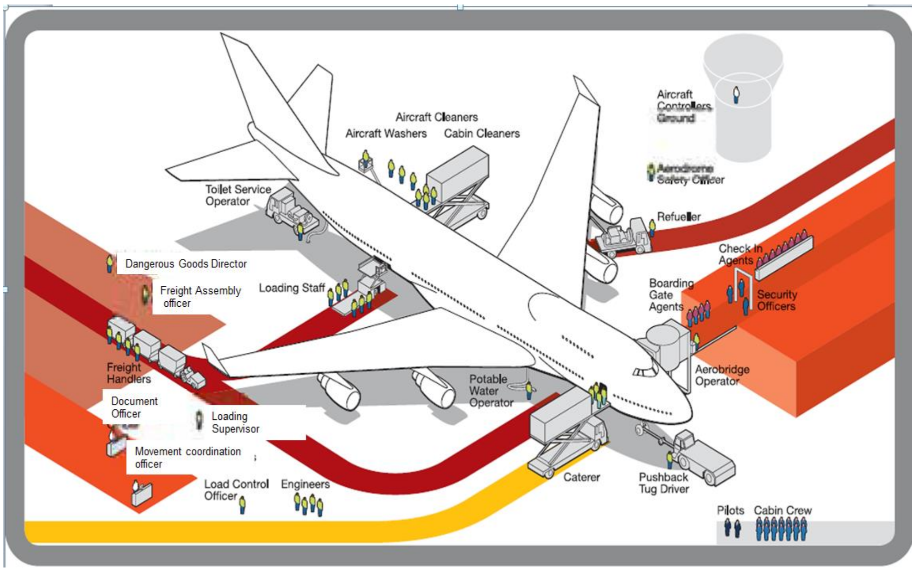


Figure 1. Ramp turnaround diagram (Source: CASA, *Flight Safety Australia* 68, p.10).

The ground handling industry: There are certain characteristics of the ground handling industry that also influence safety. Beyond the complex nature of the work itself, three factors are routinely identified as further complicating safety: a largely outsourced and subcontracted labour force, complex regulatory arrangements and the necessity to operate within an extremely cost-sensitive or restrictive economic environment.

Outsourcing and subcontracted labour: Ground services are often provided under contract to airlines by way of a local and sometimes transient workforce. IATA conservatively estimates that airlines outsource more than 50 per cent of the ground handling that takes place at the world's airports (CAPA Centre for Aviation 2014). While large airlines may provide for their own airport ground operations, especially in their local ports (i.e. aircraft loading, baggage handling, aircraft towing and pushback), in locales where the airline has a limited presence these services are usually provided under contract by other carriers or specialised ground handling organisations. The ground handling organisations may in turn engage with several different labour hire companies to provide personnel. Matthews (2012) has said that the complicated array of employment relationships on the ramp poses its own problems, such as language barriers and other communication breakdowns between the local agent and employees of multiple teams. Matthews added that unlike flight crews – which, by comparison are well-paid, well-trained, career professionals – ramp crews are relatively low-paid, receive little training and experience high turnover of employees. These factors can all influence ramp-related staff motivation, safety culture, training effectiveness and human performance outcomes.

Complex regulatory jurisdictions: Regulation is also a complex issue that can negatively influence safety outcomes in aviation ground handling. Balk (2008) explored the complexities of such arrangements in a review of how responsibility for safety is shared among and between airlines, airports and service providers. Airlines have a responsibility to ensure the safety of their aircraft and personnel when an aircraft is on the ground. Airports are also partially responsible for safe operations in and around the aircraft during ground handling activities. Airports worldwide must comply with regulations set out in ICAO Annex 14 with regard to airport facilities. Safety oversight is often managed by the state regulator through the airline's Air Operating Certificate, or even, in some cases, by the airport authority.

The regulatory situation also varies from country to country. Contracted ground service companies and their crews are often confronted with varying airline procedures and management systems (sometimes written in languages unfamiliar to them) within the same airport, which can result in confusion of interpretation or trouble deciding which rules and procedures apply. A GAO report to Congress in 2007 found that 'no federal or industry-wide standards apply to ramp operations'. The absence of a coherent regulatory structure can result in gaps in safety oversight and make difficult work of holding airports and regulators accountable for safety management. Typical outcomes are unclear or overlapping lines of safety management responsibility, uncertainty about how best to influence safety outcomes and a lack of clarity about who is responsible for initiating safety improvements. The situation is further compounded when there are unclear assignments of responsibility for the oversight of safety that would otherwise force the aviation industry to account for its performance.

Restrictive economic climate: Ground handlers often operate on narrow profit margins, as the industry is under constant pressure from increased competition and the price sensitivity of low cost carriers (CAPA Centre for Aviation 2014). Balk and Bossenbroek's (2010) study into the perceptions of ramp staff and management reported that those surveyed believed competitive pricing for ground handling services had led to postponement of investments in equipment and personnel.

When these considerable complexities and performance pressures are taken into account, it is perhaps not surprising that ground handling has not enjoyed the rise in safety improvement numbers experienced elsewhere in the industry, suggesting that new approaches are necessary.

2.3.2 The need for new research in ground safety

Each of the indicators and studies mentioned above suggests that the cost to the aviation industry (both to its people and its assets) is significant and sufficient to justify urgent research and investment in matters related to ramp safety, with Marchbank (2009) concluding that 'the cost to industry, both "human and hull" is considerable and should have already prompted investigation and investment' (p. 10). However, a review of published literature on the topic suggests that – to date – there has been a lack of investigation and research into this issue.

FSF Vice President Robert Vandel, speaking at a 2004 ground safety prevention conference, argued: 'We cannot be certain that we are fixing the correct problem unless we know what is happening and obtain an insight into why it is occurring' (p. 9). Australia's own civil aviation authority has voiced its agreement that ground handling is one area of aviation safety that has been neglected, stating: 'Despite the high reliance on human performance, ground operations have not to date served as a rich area of study for human factors research' (2011, p. 9). Most human factors research has focused on piloting errors in-flight, leading several authors to comment on the apparent gap. Piotrowicz, Edkins and Pfister (2002) have suggested that 'research on ramp safety has been scarce and human factors research even more so' speculating that this is because the aviation human factors community has been heavily focused on what they term the 'highly skilled and technical roles of pilots and engineers' (p. 2). The ATSB (2010) has also noted the paucity of research, which it sees as being due to 'a significant focus… since the 1970s, on developing risk controls for pilots and air traffic services [with sparse] attention on risk controls to improve safety in ground operations' (p. 3). Balk and Bossenbroek's (2010) recent study of ramp operations at Schiphol airport was able to identify only a handful of human factors studies from the past two decades relevant to the ramp environment.

Of these, three were in the area of training (Aircraft Dispatch and Maintenance Safety (ADAMS) (1999); Safety Training for the Aircraft Maintenance Industry (STAMINA) Safety Course for Airport Ramp Functions (SCARF) (McDonald et al, 1997)) and two were in the area of incident investigation and classification (Human Factors Analysis and Classification System (HFACS) (Shappell and Wiegmann, 2000]; Maintenance Error Decision Aid (MEDA) and Ramp Error Decision Aid (REDA) (Boeing) (Balk and Bossenbroek 2010, p. 14)). Balk and Bossenbroek (2010) suggested that human factors have yet to reach the ramp, reporting that 'human factors good practices are hardly introduced in the ground handling process [which then results in] missed opportunities to prevent incidents and accidents and to improve the safety of ramp personnel' (p. 7).

The ATSB suggested that this lack of attention is one reason for the lack of improvement in the frequency of ground operations events in Australia over the last few decades (2010). Boe, Koncak, Habekotte and Hilten (2012) noted that ground services have been targeted as an area for improvement, though they concluded that 'existing safety awareness programs have had limited effect' (p. 1).

In view of the lack of research to date, it is suggested that the time has come to consider ways in which human factors in aviation ground safety can be better understood – with a view to developing more evidence-based, future-focused prevention strategies.

Before considering the most appropriate method of data collection, however, it is first necessary to consider the benefits and limitations of the approaches currently available for this purpose. What data is collected will depend on how we believe safety is created or maintained. Concepts of safety and how it functions have changed over time. Chapter 3 explores the prevailing models of safety and accident causation that have influenced aviation and ground safety today. Chapter 4 will consider safety data collection methods available to monitor and measure safety. The objective will be to identify which models should shape our understanding of how ground safety functions and how such a model might be used to provide a framework for data collection. The ultimate aim will be to develop a new data collection tool that provides new insights into ground safety and new avenues for ground safety improvements.

Chapter 3: Accidents and Models

The previous chapter outlined the problem of safety in ground operations and the need for new approaches to improve our understanding and inform effective improvement interventions. However, before embarking on new paths, it is important to consider what approaches are already available and how they have influenced ground safety to date.

Preventing accidents requires an understanding of the causes of accidents. Over the last century there have been many attempts to model safety and to develop predictive theories of accident causation. Although none have been universally accepted, each has offered new perspectives on how safety might be created and maintained. This chapter will review some of the more dominant safety and accident causation models. Although not an exhaustive review, the aim will be to consider the prevailing concepts offered by modern safety science, and those that have been particularly influential in aviation. The aim is to identify the best approach to increase understanding and improve safety in aviation ground operations.

3.1 Models and Metaphors of Accident Causation

The field of safety science began with the aim of accident prevention (Swuste et al. 2010). Models and metaphors have frequently been used by accident investigators to try and understand why an accident has occurred and prevent reoccurrence. MacDonald (2006) argues that models are fundamental 'to both understanding how a system works and being able to devise intervention[s]' (p. 5).

A model, as it is referred to here, is a simplified representation of the salient features of some target situations. As Reason, Hollnagel and Paries (2006) have suggested, a model is 'a convenient way of referring to the shared sets of axioms, assumptions, beliefs and facts about a phenomenon that makes it possible to form an understanding of what goes on and to make predictions about what will happen' (p. 14).

Models are used to illustrate how safety is maintained or, conversely, how accidents are able to occur. Models are frequently used to represent theories (Frigg, Roman, Hartmann and Stephan 2006). They can also be used retrospectively, to explain phenomena that have already occurred, or prospectively, as predictors of potential future outcomes. Models are, therefore, useful in providing examples for the purpose of agreeing on common terms, collecting or sharing data and structuring both analyses and explanations related to how safety functions.

The wide range of models available suggests that there are many different ways of thinking about safety, and begs the question of how we can know which, if any, are accurate. Gibson (2010, p. 6) reminds us of the expression that 'all models are wrong, but some are useful'; some such models will be explored in this chapter. Most models make use of metaphors to represent the concepts in the model. The next section explores some of these models and metaphors in more detail, and considers their influence on aviation to date and potential relevance to the current research.

3.1.1 Accident chains and dominoes

Swuste et al. (2010) have described William Heinrich as providing the field of safety science with some of its most powerful metaphors. Heinrich's domino model (1927) is typical of accident models that depict accidents as a sequence of events. Heinrich likens an accident occurrence to rows of dominoes falling, where the activation of a single tile can cause all the others to fall in a chain reaction:

> The occurrence of a preventable injury is the natural culmination of a series of events or circumstances, which invariably occur in a fixed and logical order. One is dependent on another and one follows because of another, thus constituting a sequence that may be compared with a row of dominoes placed on and in such alignment in relation to one another that the fall of the first domino precipitates the fall of the entire row (Heinrich 1931, p. 14).

This implies that, in most cases, a single cause or 'unsafe act' leads to eventual or unavoidable injury (see Figure 2).

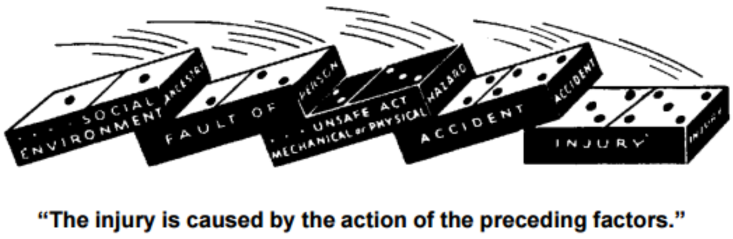


Figure 2: Heinrich's domino theory of accident causation (Cranswick, 2010, p. 37).

Heinrich's theories have evolved into contemporary accident theory, with new interpretations of the domino model that also consider the role of management (Bird and Loftus 1974, Bird and Germain 1985) and of personal and job factors (Vincoli 1994), for example. It is recognised that accidents are often the result of more complex interactions than the linear sequence the domino model suggests (Hollnagel 2010). Nevertheless, the domino model was influential in thinking about accidents as a preventable culmination of events. Perhaps even more influential, however, was Heinrich's pyramid model, which is discussed in more detail below.

3.1.2 The pyramid model

Heinrich established the pyramid model based on his study of incident data. Heinrich discovered a ratio of minor to major accidents within the data. The pyramid model is also sometimes referred to as the 'iceberg' or 'safety triangle'. Swuste et al. (2010) describe the ratio in the model as follows: 'The iceberg claims that scenarios leading to a major injury are similar to ones leading to accidents with no injury' (p. 1014).

Heinrich (1931) based this concept on a statistical analysis of accident registers attributed to *Travelers Insurance Company* – having reviewed 12,000 accident claims and 63,000 injury and illness records (which had been submitted by plant owners) – finally postulating that for every major injury there were 29 accidents that caused only minor injuries, and 300 accidents that caused no injuries at all. An illustration of this concept is included here (Figure 3).

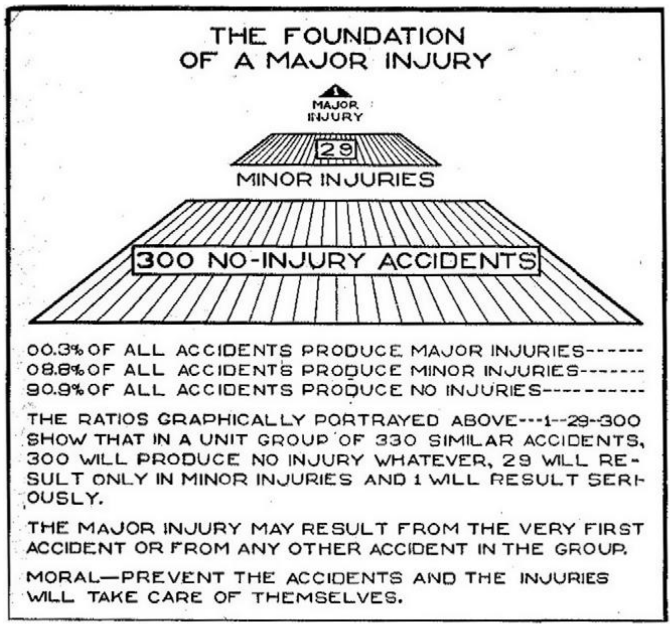


Figure 3: Heinrich's illustrated pyramid, or iceberg, of injuries (Heinrich, 1950, p. 24).

Heinrich further concluded that 88 per cent of all industrial accidents were primarily caused by unsafe acts (p. 19). It was Heinrich's belief that, since many accidents shared common root causes, addressing the incidents at the bottom of the pyramid would help prevent major accidents from occurring at the top.

Manuele (2002) argued that Heinrich's evidence for the common cause hypothesis was weak, since the data and classifications used by Heinrich to develop his theory have never been published or peer reviewed. Manuele (2002, 2003) was also concerned that the theory was derived from supervisors' accident reports from the 1920s, which may have been biased towards blaming workers rather than considering management or supervisory failures, and that this bias may account for the high number of worker-attributed unsafe acts.

However, other studies have found some support for the theory. Bird (1969), based on Heinrich's findings, tested the ratio through a study that analysed nearly two million accidents in almost 300 companies. The results confirmed the presence of a similar ratio: for every major injury (resulting in death, disability, medical complications or lost time) there were likely to be 10 minor injuries (requiring first aid) and 30 property damage accidents, suggesting an incident severity ratio of 1:10:30.

Wright and van der Schaaf (2004) also found some support for the idea that minor events and major events have common causes, also known as the 'common cause hypothesis' outlined by Peterson (1970). They proposed their own test of the hypothesis, based on data collected from a confidential report database in the UK rail industry. This data was analysed for causal factors rather than just the frequency of accident severity. Wright and van der Schaaf (2004) argued that the end results 'provide qualified support for the common-cause hypothesis within the railway domain' (p. 109).

Other studies, however, have not been able to replicate these ratio relationships in the data to support the pyramid model. Examinations of similar data sets have not produced the relationships ratio between major and minor events described in Heinrich's pyramid. For example, data from accidents in Offshore Helicopter Transportation reported by Nascimento, Majumdar and Washington (2013) unveiled 'sudden failures that cannot be reliably anticipated from more minor events' (p. 145). This lead the theorists to question the culture of reporting all minor safety issues in oil and gas as a means of preventing major accidents (Macleod 2012).

Whilst the same pattern may not be found in all datasets, it is probably true that the causes of minor incidents are often the same as the causes of more serious events. In a study of accident records from a large petrochemical refinery, for example, Laugherty (1993) found 'a great deal of similarity for both major and minor accidents', including the kinds of activities employees were engaged in prior to the accident, the sequence of events preceding the injury, the nature of the injury sustained and the parts of the body injured. Laugherty's study provides support for the notion that the conditions giving rise to major and minor injury accidents are similar (p. 273). This suggests that focusing on the causes of minor incidents to prevent major ones is still a very worthwhile endeavour.

This common cause hypothesis has influenced occupational health and safety management practices worldwide, on the basis that best predictor of what might happen is the evidence of incidents that have happened before and that, while the severity of outcomes may vary, eliminating common causes reduces the potential for incidents to occur.

The influence of Heinrich's early work can be seen in the type of data the aviation ground industry currently collects related to minor incidents, minor injuries, accidents and ground damage. There are several incident data collection programs for ground safety, managed by industry organisations such as the Flight Safety Foundations Global Information Sharing Systems, Ground Accident Prevention Program and IATA's Ground Damage Database; all help the industry to learn from minor and major events. Accident and injury data will no doubt continue to be an important source of learning from the past to prevent future harm. Heinrich's theories also encouraged the idea of thinking of accidents as a combination of actions, such as unsafe acts, and conditions in the environment.

3.1.3 The energy-damage model

Another influential theory of accident causation is that of the 'energy-damage model' (Gibson 1961, Haddon 1968). Whilst Heinrich focused on the relationships between minor and major safety events, Gibson's focus was on all the ways in which energy can cause harm. In a landmark paper, first presented in 1968 and later refined in 1972, Haddon presented a model that viewed safety hazards in terms of energy. The model provides a framework for outlining all the ways an escape of energy might occur and, if it did, how harm from this release could be minimised; according to this concept, harm occurs when energy escapes in an unplanned way (Gibson, 1961; Haddon, 1970). Road accidents, within this framework, are viewed as losses of control over kinetic energy; likewise, process accidents are attributed to a loss of containment of hazardous chemical energy, and occupational accidents are said to occur when energy escapes in the workplace to harm people – such as through falls from height or electric shocks and so forth (Kjellen 2000, p. 2).

The energy-damage model, shown in Figure 4, is based on the concept that 'damage (injury) is a result of an incident energy, whose intensity at the point of contact with the recipient, exceeds the damage threshold of the recipient' (Viner 1991, p. 42).

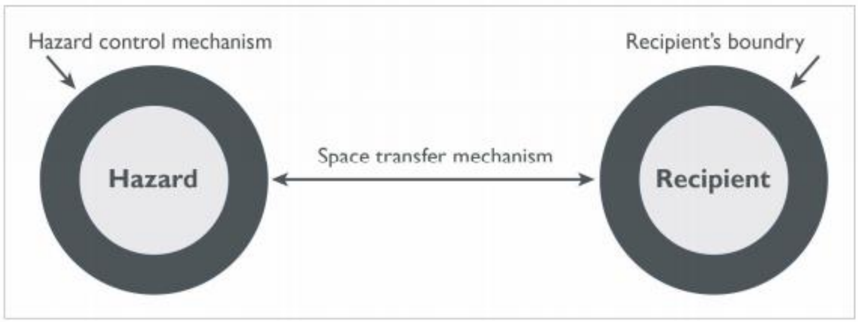


Figure 4: Energy-damage model (Viner 1991, p. 43).

Haddon proposed a matrix for conceptualising the model – essentially a table in which the columns reference factors that interact with the agent or vectors of energy. The Haddon Matrix (1970) remained largely unchanged until Runyon (1998) added a third dimension for prioritising solutions according to attributes such as cost, feasibility and likely effectiveness. Once all possible interventions have been considered, the advantages and disadvantages of each one – as a course of action – are evaluated. This allows the matrix to be actively applied to the task of assessing solutions. This energy-damage model has been used in many different settings to reduce, prevent, modify or separate energy from its corresponding targets before a loss of control can cause harm (Viner 1991; Runyan 1998). It has been influential in various industries and applications as a tool for injury prevention and for understanding accident causation (Runyan 1998; Van de Voordea et al. 2014).

The 40-plus-year application of the Haddon's Matrix caused Runyan and Baker (2009) to identify it as a successful and enduring safety concept. Others, however, view its static nature and lack of development as a problem, concluding that the Haddon concept is now out of date and has not kept pace with new developments in public health and injury prevention (Williamson and Feyer 1990, 1991). Critics point out that the model fails to reflect recent trends in safety science and human factors, or to consider the interaction between humans and energy within organisational systems. Despite widespread application, the matrix itself has not been the subject of much critical evaluation and may be overly simplistic.

However, concepts that consider the impact energy as a source of harm are still useful in that nearly all accidents can be traced back to energy displacement in one form or another. Both Gibson's Energy Model and the Haddon Matrix encourage consideration of the various ways in which harmful energy can arise. Their influence can be seen in many modern-day approaches to risk management, such as the 'hierarchy of controls' whereby people are progressively separated from harm (or energy) through six levels of controls: (1) elimination, (2) substitution, (3) isolation, (4) engineering, (5) administration and, finally (6) personal protective equipment (PPE). Level 1 is considered the most effective, and 6 least effective. It is suggested that risk management efforts should start with the preferred option of 'elimination' of the harm all together, down to the least effective control of wearing PPE.

Haddon's influence can also be found in many of the occupational health and safety practices found on the airport ramp today, which are concerned with separating ramp workers from potentially harmful energy sources such as moving aircraft, vehicles or equipment, as well as with implementing safer practices for working at heights or when transporting energy in the form of fuel or dangerous goods. Current safeguards that are intended to protect workers from harmful releases of energy include mandatory walkways, line markings, guard rails, personal protective equipment, procedures related to driving and operating machinery and rules/regulations. Separating people from the hazards will therefore continue to be an important safety strategy on the ramp.

However, the model fails to take into account the interaction between humans and the system, or the way in which humans can both create and manage risk. Adams (1995), for example, suggests that humans create and dynamically adapt to risk continuously. Conceptualising risk in terms of energy release alone ignores what Adams has suggested is the complex, social balancing act that helps regulate real human interaction with hazards.

3.2 Models of Human Performance

Whilst the energy models may have failed to consider the role of the human, developing models from the fields of cognitive psychology and human factors placed the human in a central role. Three models of human performance that have been particularly influential to safety theory and aviation safety include the Information Processing Model (Wickens 1992), The SHELL model (Edwards 1972; Hawkins 1987) and the Threat and Error Management model (Klinect, Wilhelm and Helmreich 1999). Each is discussed in more detail below.

3.2.1 The information processing model

Although not intended to be a model of accident causation per se, Wickens's (1992) Information Processing Model was nevertheless influential in evaluating the way in which human performance contributed to safety outcomes. The model attempted to explore the origins of errors and decision-making failures from a cognitive, internal-processing perspective (see Figure 5).

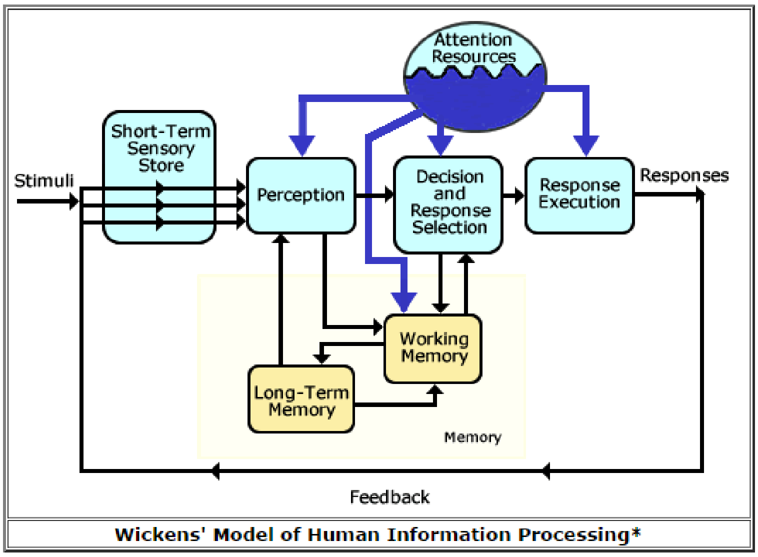


Figure 5: The cognitive information-processing model (Wickens 1992).

According to this model, errors are likely to occur when the human system is overloaded or lacks sufficient attentional capacity. The model asserts that when attentional resources are expended on one component of the information-processing system, less remains available for use by the other components. Interference or error may be expected when two or more task elements require the same processing resources at the same time. Several studies found evidence that human performance does indeed function as depicted in the model, suggesting it is an accurate reflection of cognitive processes (Wickens and Flach 1988; Wickens 2001).

The benefit of Wickens's model was that it showed the human brain to be limited in its capacity to manage information, and highlighted both the capabilities of humans and human performance limitations. Understanding these capabilities and limitations can assist the design of technology, so that it can best support human performance.

Models such as the information processing model encouraged a view of accident causation based on human error and inspired accident causation models with humans at the core. Ferrel (1997), for example, developed a theory of accidents, known as the Ferrel theory, in which accidents are caused by factors such as:

- **Overload:** where there is an incompatibility between the load and the capability of the human,
- **Incorrect response:** caused by the incompatibility of a situation in which the human is working,
- **Improper activity:** where the activity is performed improperly, either due to lack of knowledge or intentional risk-taking.

Models of information processing clearly provided new ways of thinking about safety and accidents, and the role that humans play in them. More recently, however, Dekker and Hollnagel (2004) have criticised such models, and suggested that the field of human factors has gone too far in the explanation and labelling of human contributions in accident causation.

Woods (1993) notes that it has become increasingly common in accident analyses to mistake the labels themselves for deeper insight – with accident reports attributing the cause of accidents to human failures such as ''loss of situation awareness'' or ''automation complacency''. Dekker and Hollnagel (2003) are concerned that labels are applied 'without further specification of the psychological mechanism that might possibly be responsible for the observed behaviour – much less of how such mechanism could force the sequence of events toward its eventual outcome' (p. 79).

Information processing models have undoubtedly highlighted the need to design safer systems with an appreciation of human performance limitations. However, Woods (1993) and Dekker et al. (2004) caution against labels that describe accidents as failures of 'cognition in the mind' (p. 84). Dekker and Hollnagel suggest that we should instead focus on 'cognition in the world' (2004, p. 85). Using an aviation example, Dekker and Hollnagel suggest that 'we should not be overly concerned with the performance of the pilot per se, but rather with the performance of the pilot + aircraft – in other words, the joint pilot aircraft system'.

One early example of a model that could be described as focusing of the performance of the human and the system (or 'cognition in the world') is the SHELL Model (Edwards 1972) which is discussed in more detail below.

3.2.2 The SHELL model

First developed by Edwards (1972) and then modified by Hawkins (1984), the SHELL model (see Figure 6) is a conceptual model of human interaction with systems and environments, and was one of the earliest models to depict this interaction as a simple system.

Figure 6: Hawkins's SHELL Model (ICAO circular 216-AN31 p. 6).

The SHELL model utilised a system of interfaces divided into categories of software and hardware to describe the interactions between people and technology. The model provides a framework for thinking about these interfaces in terms of how best to support the human task. It consists of five components and their related interfaces (as outlined by Hawkins 1987, p.3):

- software (i.e. policies, rules, procedures, manuals),
- hardware (i.e. equipment, tools, workspaces),
- environment (i.e. climate, temperature, vibration, noise, social factors),
- liveware, central – which places humans at the centre of activity (i.e. knowledge, skills, attitudes), and
- liveware, peripheral (i.e. teamwork, communication, leadership and behavioural norms).

The human element is the most critical and flexible component in the system, interacting directly with all other system components. However, the edges of the central human component block are varied, to represent human limitations and variations in performance. It is therefore suggested that the other system components should be adapted and matched to this human component to accommodate human limitations.

The SHELL model has been particularly influential in aviation. ICAO for example, endorses the SHELL model as a conceptual framework (ICAO Circular 216-AN31). The SHELL model has been used retrospectively to describe the failures of the Lockhart River aircraft collision with Terrain (ATSB Report 2005 019.77). CASA (2012) suggests the model 'illustrates how important it is to understand the human contribution to an accident in context, rather than simply labelling what somebody did as "operator error"' (p. 3).

The SHELL model has also been employed to analyse and develop successful interventions in ground safety. For example, Drury et al. (1998) used the SHELL model components to highlight potential intervention strategies for the airport ramp, such as a liveware/liveware training – which targeted first-line supervisors – and a liveware/software intervention – which targeted changes in procedures. Where applied, such concepts have improved the design of interfaces and tasks on the ramp to better suit the capacities and limitations of human beings (Edwards 1972; Hawkins 1993).

Although the model clearly provided new ways of thinking about human interaction with the environment, it has limitations to its explanatory power. Eurocontrol (2000) describe the model as overly simplistic, noting that the building block diagram does not cover the interfaces that are exterior to human factors (e.g., interactions such as hardware-hardware; hardware-environment; software-hardware), and is only intended as a basic aid to understanding.

Another limitation of the model is that it does not describe human performance dynamically, instead providing a static representation of the human within the environment. It cannot, for example, account for the ways in which human beings manage threats, change within the environment or adapt their performance to manage risks.

Nearly a decade later, a more dynamic model would emerge that attempted to explore how humans actively manage risks in their environments; this is known as the Threat and Error Management model.

3.2.3 The Threat and Error Management (TEM) model

The TEM model, developed by Klinect, Wilhelm and Helmreich (1999), offered a view of human interaction with the environment. The FAA (2006) has described the TEM model as 'a conceptual framework for understanding operational performance in complex environments' (Appendix 1, p. 1). The model was originally created to capture flight crews' performances in commercial aviation; however, the FAA (2006) explains that the model is 'generic and can be applied to numerous work situations' (Appendix 1, p. 1) The advantage of the model was that it was able to focus 'simultaneously on the operating environment and the humans working in that environment' (FAA 2006, Appendix 1, p. 1).

The model attempts to portray how humans deal with threats in the environment (including interfaces with people, equipment or systems) and to describe how humans manage those interactions, either successfully or unsuccessfully (see Figure 7).

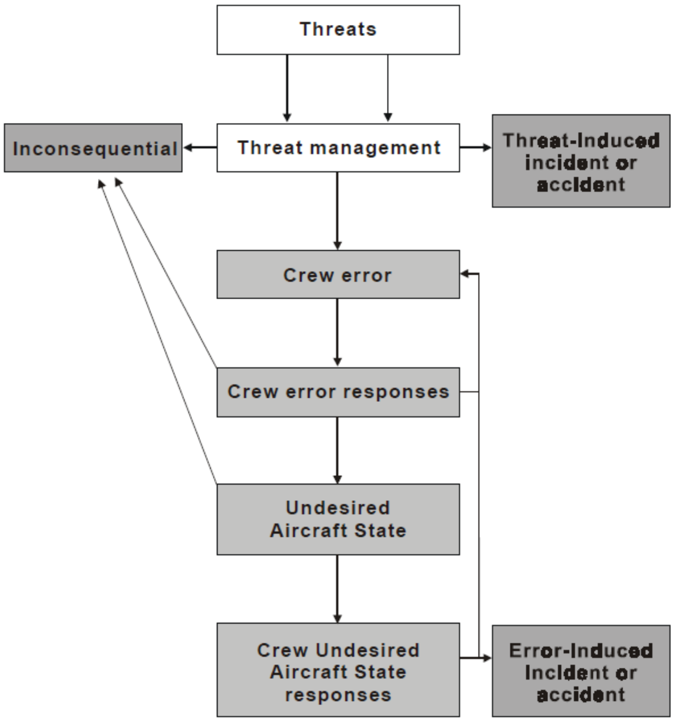


Figure 7: The TEM model of human-interface interaction (ICAO document 9803 2002, Appendix 2.2).

In TEM, *threats* are considered situations that must be managed by the cockpit crew during normal, everyday flights. Threats increase the operational complexity of a flight and thereby pose a safety risk. *Errors* are described as actions or inactions that lead to variations in crew and/or organisational intentions or expectations. According to the TEM model, threats and errors can go undetected, be effectively managed or result in additional errors that require subsequent detection and response. If crew response is ineffective, the consequence of threats and errors may result in undesirable aircraft conditions – or undesirable aircraft state (UAS) – which poses an increased risk to safety. This might include, for example, an unstable approach or long landing that will likely result in an incident if action is not taken.

The model makes the assertion that errors are normal and frequent, but it is the way in which they are managed by humans that contributes to a safe or unsafe outcome. Its ability to describe how errors occur, under normal operating conditions, rather than after an incident, is seen as an important development in the study of human performance.

Helmreich (2000) has further suggested that since accidents occur infrequently, it would be better to conduct an examination of errors under routine conditions to understand how humans manage everyday threats to achieve normal, safe outcomes most of the time, rather than studying the rare occasions on which they did not. Helmreich believed this would improve organisations' understanding of human performance and their ability to maximise safety margins.

The TEM model assumes that threats and errors may sometimes go undetected and that, on other occasions, they may be effectively managed. Sometimes a mismanaged external threat or error leads to a situation that represents an increased risk to safety. Recovery from any one of these states can be achieved without consequence. If for some reason, however, the situation is not managed, an accident or incident may occur. The logical conclusion drawn from the model is that the *management* of threats and errors should be the focus of an organisation's efforts to maintain safety (Klinect, Wilhelm and Helmreich 1999; Thomas 2004).

Unlike the Information Processing model, however, there are only a few studies that have investigated the interactions suggested by the model. In one example, Thomas (2004) investigated the link between a set of CRM behaviours or Non-Technical Skills (NTS) and TEM outcomes. Thomas found, for example, that the presence of some CRM behaviours such as 'communication', particularly before take-off, was a predictor of effective threat and error management. Thomas also reported that 'assertiveness' and 'vigilance' were predictors of effective error management during the time critical descent, approach and landing phase (p. 223), while skills such as 'decision making' and 'situational awareness' were found to be related to effective threat and error management in all phases of flight. Thomas also found that, contrary to what might be expected from the model, higher numbers of threats were linked to improved threat management. Thomas also refers to some ‘contextual factors’, such ‘operational pressures’ and late departures, linked to poor error management.

Thomas’s study concluded that some non-technical skills and contextual factors are predictors of threat and error management and therefore suggests that including these skills in training would ‘minimise risk’ and ‘enhance operational performance’ (Thomas 2004, p. 229). The research by Thomas does not, however, attempt to either validate the TEM model or question its validity. It starts from the premise that the Threat and Error Management approach proposed by Klinect, Wilhelm and Helmreich (1999) ‘form[s] fundamental causal components of incidents and accidents’ (p. 208). Thomas therefore argues ‘that the management of threat and error must form the focus of any organizations attempts to effectively maintain safety in high risk operations’ (Thomas 2004, p. 208).

Thomas’s assumed validity of the model and its causal relationship between accidents and incidents is perhaps premature. Helmreich et al. (1999) merely describe how the TEM model was developed to help organise and make sense of data collected from observation field studies (Helmreich, Klinect and Wilhelm 1999; Helmreich 2002). Whilst the model may offer intuitive explanations for how human beings interact with risk, its creators do not claim that the model itself or its theoretical underpinnings have been validated. It is offered merely as a model of human performance developed from their body of work. Nevertheless, the model has been widely applied in aviation in a range of circumstances, including as a framework for classifying human behaviour, as a training tool and as an accident investigation tool (ICAO 2002; FAA 2006; IATA 2014). This has led to criticism from some theorists that the model has been applied without sufficient scrutiny, or even that it is invalid. (Dekker 2001, 2003; Hollnagel and Amalberti 2001; Hollnagel 2009).

Despite these criticisms, the TEM concept has continued to proliferate in aviation including ground safety. For example, TEM is recommended by the industry body European Commercial

Aviation Safety Team (ECAST) as an accepted tool to help ramp workers manage their own errors, although it is not clear upon what evidence this recommendation is made. At first glance the TEM model appears to offer a useful model exploring human and safety performance that is already accepted in aviation. However, the concerns regarding its validity as a model require further scrutiny. These issues will be explored in more depth in later chapters.

3.3 System Models

Whereas human performance models focus on how humans interact with the system, system models focus on the functioning of the system itself. The organisational perspective in safety and accident investigation can be dated from Landau (1969), Turner (1978) and Perrow (1984), with Perrow introducing the idea that it is the modern organisational system that makes accidents inevitable, where hidden interactions create unpredictable events. He suggested that multiple unexpected failures are built into complex and tightly coupled systems. These ideas can be seen throughout the development of the system models of accident causation, which have been influential in safety management in general and in aviation in particular.

Kjellen (1995) was one of the first to model the organisation as a system through the use of deviation models. In this model of safety, variations are similar to *unsafe acts* performed or initiated by people, but also include *unsafe conditions* brought about by, or in relation to, machines or the environment. Deviations, in this model, are acts or conditions that can lead to accidents and fall outside the norm of what is planned, usual and accepted. In practical terms, the model suggests that, by collecting data on deviations noted in safety inspections and near-accident reporting or checklists, it is possible to identify the circumstances that lead to accidents and prevent them from reoccurring (Kjellen and Hoveden 1993). The deviation model is thought to have inspired the modern risk-assessment model, which Stellman (1998) has said focuses on precursors as early warnings of accidents.

More complex systems models have evolved that attempt to take into account the complex social interaction between people, technology and organisations known as ‘socio-technical systems (Trist 1981; Fahlbruch and Wilpert 1999). These models have developed from the earlier concepts of Heinrich (1931), where accidents are rarely viewed as occurring as the result of a singular human failure or sequential chains of events, but rather as the result of a complex set of interactions. To make sense of those interactions, organisational accident researchers sought to find a way of describing organisations and establishing guidelines for how they functioned as sociotechnical systems. One of the most enduring models for describing this has been Reason’s ‘Swiss Cheese’ Model.

3.3.1 Reason’s ‘Swiss Cheese’ model

The Reason model heavily informs popular thought related to safety and accident causation today (Perneger 2005) and has been described as ‘undoubtedly the most popular accident causation model’ by Underwood and Waterson (2014, p.76). It has been widely applied in a number of different industries, including healthcare (Stein and Heiss 2015), construction (Hosseinian et al. 2012) process industries (Al-Shaninia 2014), the energy sector (Lorenz and Ziff 2001) and the transport sector (Underwood and Waterson 2014).

Hudson (2003) describes how the Swiss Cheese model was originally developed as part of a research programme in a major oil company. Reason’s 1990 model of organisational-accident causation built upon the idea of unsafe acts/conditions, Haddon’s energy model and other organisational models to depict the way in which unsafe acts or conditions might line up with holes in the organisation’s defences, allowing an organisation’s accident trajectory to occur – not unlike Haddon’s release-of-energy theory. It is sometimes referred to as the ‘Swiss Cheese’ model because harmful influences are depicted as passing through a series of gaps or holes, similar to layers of Swiss cheese, which appear in an organisation’s defences (Figure 8).

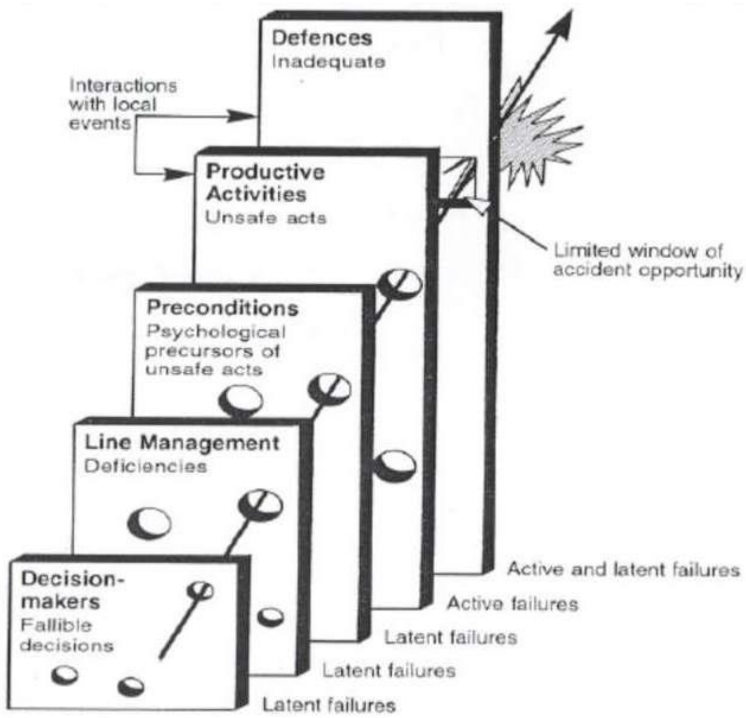


Figure 8: Reason's 'Swiss Cheese' model
(Crew Resource Management Aviation Safety, 2011).

The trajectory looks similar to a release of energy in the model suggested by Haddon; however, Reason's model suggests that failures in an organisation's defences are what allow accident potential to continue on its path. Reason (1990) described these failures as being either 'active' or 'latent'; active failures were associated with front-line operators of the system – the results of which he argued were almost always immediately observable – while latent failures were thought to lay dormant within the system for a long period of time. If, in fact, it is these organisational deficiencies that create the necessary conditions for accidents to occur, then the model could offer new opportunities to defend against accidents. The Reason model emphasised the importance of managing latent organisational conditions; if these could be addressed, the model could be used proactively to prevent undesirable events from occurring.

Reason's Swiss Cheese model is extremely influential in aviation today (Salmon et al. 2012). Hudson (2003) describes how it was 'rapidly taken up by the world's aviation industry'

(p. 10) soon after the Australian Bureau of Air Safety Investigation (BASI) used it as the basis for all its major reports.

In 2008, accident investigators at the ATSB amended the Reason model to address a number of concerns (Walker and Bills 2008). Shorrock, Young and Faulkner (2005), for example, criticised the assumption that the causes of accidents always involve latent conditions and management decisions further up the chain. Walker and Bills (2008) also suggest the model emphasised organisational factors over technical failures or individual actions in investigations (p. 35). Walker and Bills were also keen to develop a model that showed an organisations 'normal system performance' as well as the potential for 'deviations' that can lead to accidents (Walker and Bills 2008, p. 36) In their adaption of the model, Walker and Bills depict the organisation as achieving its production goals most of the time (normal operations) whilst allowing the accident trajectory to occur only when the organisation's defences, or risk controls, are inadequate, as shown in Figure 9 below.

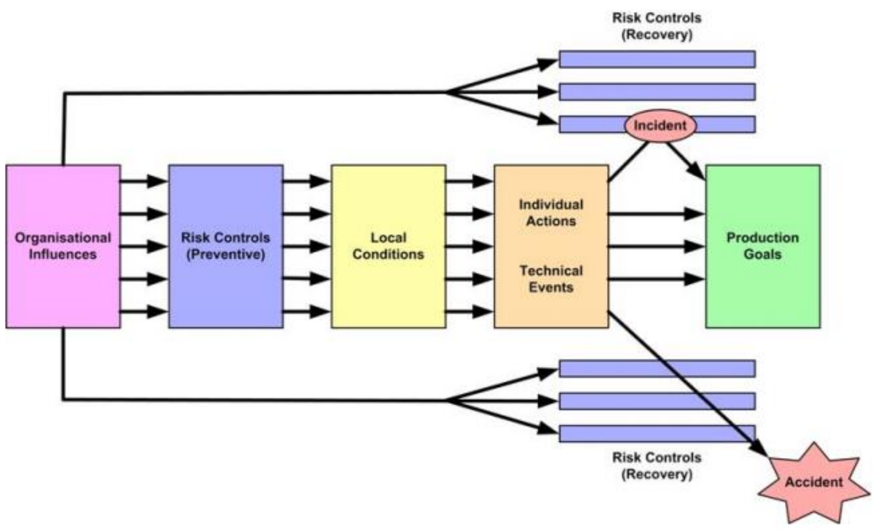


Figure 9: ATSB adaption of the Reason model (Walker and Bills 2008 p. 36).

Variations of the Reason model are now firmly embodied in aviation safety concepts and accident investigation tools (Li, Harris and Yu 2008). The Reason model has been applied to accident investigation in many areas of aviation, including ground operations (Walker 2008), where it helps to conceptualise how ramp accidents can occur. Still, the model is not without its critics. Some researchers have commented on the model's lack of specificity about a number of its features, such as how the holes in the layers of cheese line up (Le Coze 2013; Wiegmann and Shappell 2003). Shorrock et al. (2015) are concerned that the model depicts individual actions as symptomatic of organisational failures, whereas they suggest ‘sometimes, the actions or activities of human beings may not be symptomatic of organisational failures at all; they are just the localised errors of individuals’ (p. 392). They dispute the assumption that all human errors can be controlled, suggesting that ‘better defences could help but some errors would overcome even well-planned and maintained defences’ (p. 395).

The Reason model in its original form may overemphasise organisational failures and the role of defences, but it proved to be a step forward in thinking about accident causation as systemic involving a complex interplay of factors. It has assisted organisations to consider additional ways accident trajectories might develop, as well as the strength and depth of current defences to prevent harm. The Reason model has changed the way practitioners think about safety and accident causation enormously and has been particularly influential in aviation particularly for modelling accident investigation (Walker and Bills 2008).

3.3.2 Safety management system models

The Reason model suggested that the root causes of accidents are usually composed of many complex, interrelated factors within an organisation. Brown et al. (2000) went further, arguing that behind every accident is a failed organisation. Hale et al. (1997) note the development of this idea from a series of reports into major disasters (e.g. Flixborough, Department of Employment, 1975; Zeebrugge, Department of Transport, 1987; Kings Cross, Department of Transport, 1988; Clapham Junction, Hidden, 1989; Piper Alpha, Department of Energy, 1990), in which the main emphasis was on the failings of management to ensure activities were designed, operated and maintained with sufficient safety. The idea that organisation and management factors were the key to keeping an operation safe, also gained significant ground in the aviation industry (Santos-Reyes and Beard 2002).

The model of a safety management system, or SMS, evolved as a way of describing the organisation and management factors thought to be important to produce safe outcomes. With its roots in a manufacturing tool known as Total Quality Management (TQM), SMS was predicated on the idea that feedback loops can provide a mechanism for continual adjustment and improvement in any process, including safety (Hale 1985; Deming 1990). SMS models highlighted the need to assess aspects such as management structures, processes for planning, procedures, auditing and reviews, as shown in Figure 10.

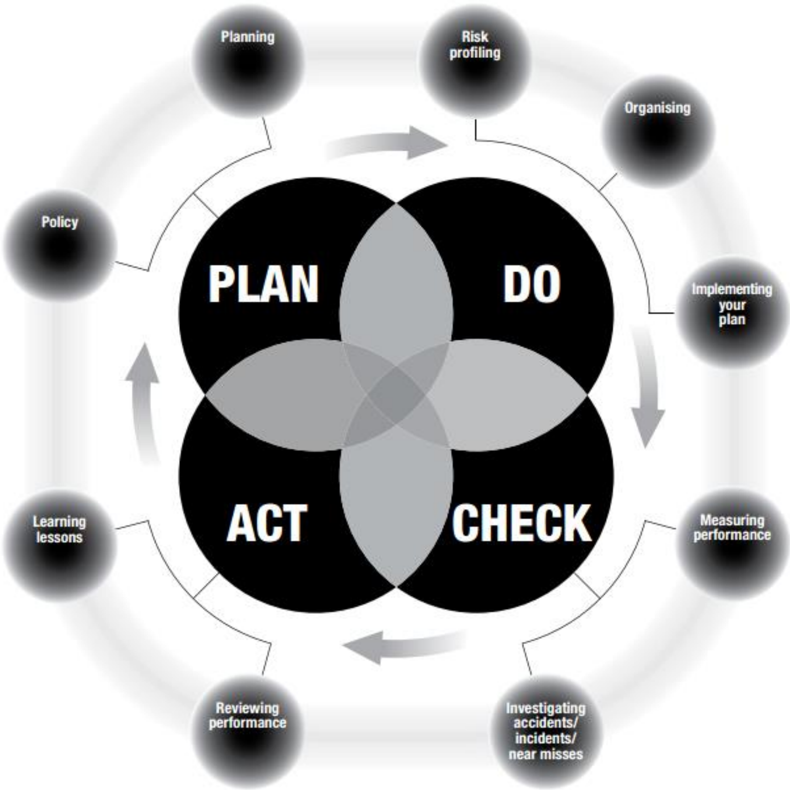


Figure 10: HSE Plan Do Check SMS model, from the UK Health and Safety Executive SMS guidance (HSG65 1991, 2013).

The SMS idea was wholeheartedly adopted by the aviation industry. In 2006, ICAO produced a manual describing the elements an SMS should have, which included safety policy and objectives, safety risk management, safety assurance processes, and safety promotion. SMSs are now already standard in air traffic services and in most areas of commercial aviation. ICAO made SMSs part of its certification requirement for aerodromes in 2005. In Australia, airlines are now required to have a safety management system that covers their ground handling operations (CASA82.5).

Dijkstra (2007) argues that publications about SMS 'do not contain specifications of how methods should operate, but merely list components that should be part of the SMS'. Furthermore, Dijkstra argues that, so far, there is not much evidence to justify the proliferation of the SMS, suggesting that 'claims regarding the effectiveness of a SMS are not yet substantiated by scientific research' (p. 1).

3.3.3 Models of safety culture

Although systems and processes are important in managing safety, new theories emerged that suggest that there were other, subtler features of an organisation that may contribute to an organisation's safety success. For example, Zohar (1980), and later Schein (1992), suggested that these features and characteristics were an inherent part of an organisation's 'safety culture'. Safety culture has been an explanatory concept to describe characteristics of organisational culture important to safety, such as the internal behavioural norms, beliefs and everyday practices for handling hazards and risks. The term 'safety culture' gained ground in the aftermath of high-profile accidents like Chernobyl, when it was used to describe a collection of qualities thought to be associated with organisations' safety outcomes.

Poor safety culture is widely believed to be a contributor to hazardous events. A widely referenced definition of culture was put forward by the Advisory Committee on the Safety of Nuclear Installations (ACSNI), as 'the product of individual and group values, attitudes, perceptions, competencies, and patterns of behaviour that determine the commitment to, and the style and proficiency of, an organisation's health and safety management' (HSC 1993, p. 23).

There has been a wide-ranging discussion in the academic literature in relation to what safety culture is, what features of an organisation contribute to it and how it might function to influence safety. Guldenmund (2002) argued that, despite decades of empirical research, little consensus has been reached on the different aspects commonly associated with the concept (p. 215). Several different models have been put forward to explain how culture might function in an organisation (e.g., Flemming 2000; Clarke 2000; Hudson 2003; Reason 1998; Cooper 2000; Cooper 2002).

In a review of safety culture models and their use in industry, Heese (2012) described the variety of safety culture models, which have been developed as falling into two broad categories:

> 1) those that refer to concepts of safety culture maturity (Flemming et al. 1996 and Hudson 2003), or
>
> 2) those that refer to the dimensions underlying safety culture elements (Cooper 2000; Mearns, Kirwan and Kennedy 2009; Reason 1997; Wiegmann, Zhang, Von Thaden, Sharma and Mitchel 2002).

One of the more influential models of culture and accident causation in aviation has been the Cultural Maturity Model, originally proposed by Westrum and developed further by Hudson, as shown in Figure 11 below. The model suggests that by measuring where an organisation is on the spectrum, steps can be taken to achieve the next level of safety maturity. It is assumed that models at the mature end of the spectrum would produce positive safety outcomes, although this has not been independently verified.

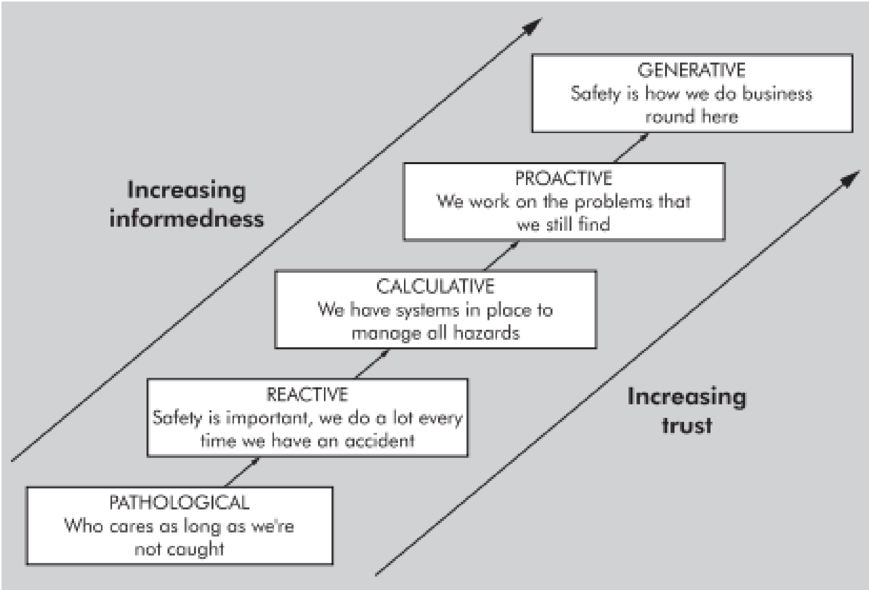


Figure 11: Safety culture maturity model (Hudson 2003, p. 9).

While Hudson depicted culture as a journey towards maturity, other models focused on describing the ideal dimensions of a positive safety culture (Cooper 2000). Some models share similar common themes such as the promotion of trust, openness of communication, leadership and management commitment to safety, for example. However, several authors have commented on the lack of a unifying theoretical model in this area (Guldenmund 2000; Williamson et al. 1997; Flin et al. 2000) and evidence to validate the dimension proposed by the models has proved elusive. (Dedobbeleer and Beland 1991, Coyle et al. 1995). This is in part because it is difficult to draw direct comparisons between dimension labels used in each study (Flin et al. 2000) and attempts to relate aspects of particular models to their corresponding dimensions has proved difficult (Lopez et al. 2013).

Despite the difficulties in validating the models or explaining how these dimensions relate to safer outcomes, the idea that safe cultures create safe outcomes is still particularly influential in aviation safety today (Mearns et al. 2013; Bennet, Hellier and Weyman 2015; Halford 2016). Two models that Le Coze (2015) describes as ‘central contributions’ to the field of safety over the past decade, including in aviation, are the High Reliability Organisation

(HRO) model and the Resilience Engineering concept. These models are discussed in more detail below.

3.3.4 High Reliability Organisation (HRO) model

Rather than try to define the features of an organisation that may (or may not) be related to safety outcomes, an alternative approach has been to find organisations with excellent safety records, known as high reliability organisations, then try to establish what cultural features and characteristics they have in common. The High Reliability Organisation (HRO) model began as a project to define the characteristics of organisations with excellent safety records. Through a series of empirical studies, HRO researchers have attempted to identify features common to organisations operating in high-hazard industries but demonstrating low numbers of safety incidents or accidents (Roberts 1990; LaPorte et al. 1991; Rochlin 1993; Weick et al. 1999; Weick and Sutcliffe 2001).

LaPorte, Rochlin and Roberts (1991) describe high reliability organisations (HROs) as those that rely on technology with inherent risk and the potential for human error, where the possible consequences of failure are precluded learning through experimentation. Roberts (1990) proposed that HROs are those that, despite working in risky environments, have enjoyed a record of high safety over long periods of time. The HRO concept has been developed to describe organisations that, by definition, are achievers of high reliability, but also organisations that actively seek out reliability through their organisational processes and characteristics. For example, Sutcliffe and Obstfeld (1999) described HROs as embodying 'processes of mindfulness that suppresses tendencies toward inertia' (p. 81). Weick and Sutcliffe (2007) describe an HRO environment as one where all workers look for, and report, small problems or unsafe conditions before they pose a substantial risk to the organisation and when they are easy to fix. Chassin and Loeb (2013) suggest that HROs are able to recognise early indicators of threats to organisational performance, prizing themselves on ‘the identification of errors and close calls for the lessons that point to specific weaknesses in safety protocols or procedures that can be remedied to reduce the risk of future failures.’ (p. 461). Hartley (2011) describes the ultimate goal of an HRO as to identify the gap between work as imagined and work as done. Hartley (2011) suggests that by understanding how and why this gap occurs an HRO can ‘align, tighten and sustain’ better performance (p. 13).

Models of organisational systems have been extremely influential in aviation, with HROs held up as an example of best practice (Ciavarelli 2005). Commercial airlines are sometimes described as HRO organisations, with airplane accidents occurring at a rate of less than one major accident for every 3.1 million flights (IATA 2016). However, in the U.S., for example, commercial airlines also have nearly twice the national average of fatal workplace injuries. Marx (2017) argues that this suggests there is not something unique about the characteristics of airline organisations, or they would have also improved safety in other areas of their operation. Rather, Marx argues that ‘all organizations will be reliable only around those things they value by putting in the time and resources to achieve an extraordinary level of reliability’ (2017, p. 1).

This cast doubt on the notion than an organisation will become safer by adopting HRO approaches. Boin and Schilman (2008) argue that any claim that organisations can be made more reliable by adopting the characteristics of HROs has yet to be demonstrated. The FAA has attempted to develop tools for understanding how HROs develop (O’Neil 2012) but its own research suggests that HROs cannot be easily created and may, in fact, take many years to develop. It is still unclear whether an HRO model can be successfully imposed on an existing culture. Vincent and Amalberti (2016) note that models ‘must also be accepted by the front line operators and be consistent with the values and culture of both the team and the wider organisation’ or safety benefits are likely to be localised and modest (p. 35).

Vincent and Amalberti caution against the idea that there is a 'one size fits all' model for safety as suggested by the HRO model and suggest that different environments may suit different approaches. Le Coze (2016) argues that parallels with HRO theories and models can also be found in the development of models of Organisational Resilience, which is discussed further below. High Reliability Organisations are by definition, resilient to risks and perturbations in the systems that could lead to accidents. Some theorists have focused on the quality of resilience itself, as the defining characteristic of what keeps an organisation safe. Lekka 's (2011) literature review of HROs outlined two of the most defining characteristics of HROs as being able to 'anticipate problems' and 'contain unexpected events' (p. 5) , two qualities that have become synonymous with models of Resilient Organisations.

3.3.5 Models of organisational resilience

Resilience refers to an individual's or organisation's capacity to respond and cope with unpredictable changes, risks and perturbations within the system. Like HROs, models of resilience encourage organisations to develop attributes necessary to deal with uncertainty and adversity (Norhouse 2000).

Eurocontrol (2009) describe resilience as 'the intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances, so that it can sustain required operations under both expected and unexpected conditions' (p. 2). Hale and Heijer (2006) defined resilience as 'the ability, in difficult conditions, to stay within the safe envelope and avoid accidents' (p. 40). They also identified areas of overlap between resilience engineering and characteristics that present themselves in high-reliability organisations and safety cultures.

Resilience models attempt to describe the properties an organisation needs to respond flexibly and dynamically to risk whilst maintaining safety margins (Rasmussen and

Link 1981; Woods 1988; Hollnagel 1993; Reason 2008; Woods and Cooke 2003; Dekker 2002, 2004). Le Coze (2008) has described it as having ‘rapidly emerged as an approach to be reckoned with in relation to how we understand and manage safety and vulnerability in socio-technical systems’ (p. 1).

Resilience models attempt to go beyond the Swiss Cheese analogy, to deal with non-linear complex events that the Reason model was unable predict, where disproportionately large consequences can arise from seemingly small variations in performance and conditions’ (Eurocontrol 2009, p. 5). Several conceptual metaphors have been used to describe resilient and non-resilient organisations. Brand and Jax (2007), as well as Longstaff, Koslowski and Geoghegan (2013), have likened resilience to a spring or other material possessing elasticity. Hollnagel and Woods (2005), Carvalho (2011) and Lay and Matthieu (2014) have characterised resilience to a resonant ability to rebound. By contrast, non-resilient organisations are described as being rigid, fixed or brittle. A resilient organisation is, therefore, thought to be one that remains in control by being able to adapt and rebound in the face of everyday safety challenges.

Resilience models also stressed the adaptability of humans to respond to risk. This was in contrast to previous models such as the Reason model, where humans create ‘unsafe acts’ that appear as holes in an organisation’s defences, or HRA analysis where human behaviours are regarded as opportunities for risk. In resilience models, humans are seen rather as proactive agents who respond to risk in flexible and adaptive ways (Woods and Hollnagel 2006). Resilience theorists also discuss the need to understand the gap between ‘Work as imagined vs work as actually done’ – (Hollnagel, Woods and Levenson 2007, p. 86), recognising that actual work drifts from prescribed protocol and procedure. Hollnagel argues that behaviours seen as non-compliance may be viewed from those involved as a mark of professionalism for getting the job done despite constantly changing operational circumstances, which they describe as the 'mark of a resilient organisation' (p. 87).

According to Le Coze (2016), the central contention of resilience is that it is more efficient to concentrate on the ability of individuals to cope with complexity rather than on their errors. Hovden et al. (2008) agree that resilience models allow us to explore the adaptive capacities of humans rather than how they might fail. Rollenhagen and Reiman (2011) argue that rather than studying humans as a cause of failure we should study the ways in which people creatively manage risk, by fixing defects in technology, compensating for bad tool or work design, or stopping a dangerous chain of events based on intuition (2011, p. 15).

Therefore, to develop a resilient organisation it is first necessary to recognise that safety and production are intrinsically linked. Systems are typically underspecified and not all risks can be planned for in procedures. Practically, to get anything done, humans must constantly adjust to real-world situations. Performance variability is therefore 'the reason why things go right, and sometimes also the reason why things go wrong' (Eurocontrol 2009, p. 12). The aim of Resilience Engineering is to develop models and methods to improve an organisation's ability to succeed under these varying conditions.

MacDonald (2006), however, has described the concept of resilience as a new synthesis that is in danger of 'promising much but delivering little' (p. 4). He is also critical of resilience metaphors and explanatory principles, such as 'resonance' and' emergence', asking: 'How do such metaphors map onto real organisational processes?' and, in relation to the concept of resonance, 'What rhythmic organisational process is being amplified, by what forcing factor, to create what organisational consequences?' (p. 4), suggesting that resilience concepts lack substance and rigour.

Oxstrand and Sylvander (2010) suggest that rather than offering a paradigm shift in the way safety is conceptualised, the resilience model is ‘not much more than a relabelling of well-known concepts of safety culture’ (p. 135). MacDonald (2006) agreed with its other detractors, suggesting that: ‘The attraction of resilience engineering has more to do with the weaknesses of current alternative theoretical models to come up with a convincing demonstration of their utility to help manage the complexities of the real world than [it does] with the real power of this theory’ (p. 1). Hovden (2008) described resilience models as ‘promising’, but suggested that the field is immature with regards to practical applications for industry (p. 8).

Although controversial, the concept of Resilience continues to grow in influence, particularly in the aviation industry (Eurocontrol 2009), although there may be some way to go before the Resilience models are widely accepted. Hollnagel, Paries, Woods and Weatherall (2011), for example, conducted a survey of aviation professionals to assess their readiness to embrace resilience models. The results suggested that there was ‘still work to do’ to create models and tools that aviation safety professionals would apply in their daily work (p. 273).

Nevertheless, the resilience model has offered an alternative view safety where, instead of trying to predict and control risks, there is recognition of complex, underspecified systems where all the possible risk eventualities are not foreseeable. If this is the case, we should instead seek to develop resilient systems that can respond flexibly to variabilities and that allow humans to adapt and respond to manage risk. This suggests that any method for measuring and monitoring safety should be able to take into account the ways in which humans manage risk every day, as well as where they might fail (Reason 2008).

3.4 Summary and Implications for the Current Research

The discussion has highlighted that accidents are complex events and there are many different interpretations of how accidents may arise. Models have progressed from simplistic linear models to complex systemic models, all of which have contributed to today’s prevailing safety management ideology and their influence can be seen in safety management in aviation today.

Having a model is an important first step in conceptualising how safety is created and where best to intervene. Perneger (2005) suggested applying models that encourage use of common terms and data collection/sharing techniques would also be useful. When considering which concepts to take forward in a study of safety for ground handling, it would seem that applying a model to aid understanding of how accidents occur would be a useful starting point.

Of the models that have been reviewed here, it is clear none have been universally accepted, although some have been more influential than others, particularly in aviation. The early models of Heinrich (1931) conceptualised accidents as a chain of events that could be prevented, and encouraged organisations to focus on unsafe accidents and conditions to prevent their occurrence. Haddon (1968, 1970) helped to conceptualise all the ways in which harm can arise, and the need for organisational defences.

Human performance models, like TEM, have emphasised the need to understand the capabilities and limitations of people in the system. If errors are normal and frequent, as suggested by the TEM model, then how they are managed is important for producing safe outcomes. This suggests we need to know more about how people manage everyday threats under routine conditions, not just when things go wrong.

Reason's model (1995) has highlighted the need to consider the complex interaction between humans the system. This suggests it is important to consider how the system supports the human task, and defends against the potential for error. It should be recognised, as Walker and Bills (2008) suggest, that whilst the organisation mostly achieves its goals in normal operations, deviations can lead to accidents. Considering how an organisation can deviate from normal operations helps to identify the multiple avenues for accident trajectories to develop. Reason's model also help to conceptualise the importance of risk controls and organisational defences in preventing harm from occurring.

The SMS model and management approach encourages the proactive monitoring and measuring of safety. Collecting data and creating feedback loops are essential for continual adjustment and improvement in any process, including safety. Mechanisms are needed to identify early indicators of threats to the system's performance (Loeb 2013) and provide the data and feedback that such systems require in order to drive continuous improvement.

The models of safety culture suggest that some characteristics of organisations are important to the safety outcomes it produces as suggested by the HRO models (Weick and Sutcliff 2001). The ability to adjust to perturbations and keep the operation within the safety envelope are important to organisational resilience (Hale and Heijer 2006). Although it is important to understand how and where humans in the system can create risk, it is just as important to understand how they can respond to contain and manage risk to safe outcomes most of the time. Hollnagel et al. (2007) ask: 'How can we measure this gap between the system as designed or imagined and the system as actually operated? And perhaps more importantly what can we do about it? The challenge is to make the gap visible and provide and provide a basis for learning and adaption where necessary' (p. 98).

If ground safety improvement has stalled (Verschoor and Young 2011; Passenier 2015), as discussed in Chapter 2, then new data and feedback loops are needed to inform safety management initiatives and drive continuous improvement. This review of models of accident causation suggests that the current research would benefit from a model of safety performance, to provide common terms and facilitate the classification and sharing of data. Models that help identify what should be measured, and how that data can best be organised, would be a first step to inform our understanding of safety in ground operations.

Ideally the model should consider the interaction people and systems, to produce safe outcomes most of the time, as well as the deviations from normal operations that allow accident trajectories to occur. The model should also consider how organisational defences and risk controls prevent accidents pathways from developing or causing harm. A model that can combine these ideas, and build on concepts already accepted within aviation, would be particularly beneficial. The next chapter will consider how models have influenced the way in which safety is measured and managed with a view to identifying the most appropriate model upon which to base a new method for measuring and influencing safety in ground operations.

Chapter 4: Methods for Measuring Safety

Accident models guide the choice of safety performance indicators (Hovden, Albrechtsen and Herrera 2008, p. 2) This chapter will consider the three main methods for collecting data about safety: safety incident indicators, self-report data and observation. The advantages and limitations of different data sources and collection methods will be explored. Their potential application in aviation ground safety will be discussed and, finally, the ideal features of a data collection tool for ground safety will be proposed.

4.1 Incidents, Indicators and Measures of Safety Systems

Early models of accident causation inspired the collection of a great deal of data about incidents and their precursors in order to prevent reoccurrence. The search for more proactive measures of safety has also lead to a wide range of possible safety indicators that can be measured. The Reason model has been influential in trying to measure the performance of the organisational system. The introduction of the safety management system has inspired the measurement of safety performance across a broad range of proactive and reactive measures in order to drive feedback loops and continuous improvement. These data collection methods and their use in aviation to date are discussed in more detail below

4.1.1 Accident and incident data

As early as the 1930s, organisations were collecting incident data in the hope of preventing more serious accidents (Reynard 1991). Incident data is used today both as a measure of current safety performance and to identify trends and patterns, to better understand where harm might arise and develop suitable safety management interventions (Lindberg et al. 2010).

Tarrants (1963) was one of the first to recommend expansion of related databases to include situations that nearly resulted in an accident, also known as ‘near misses’. A near miss, as defined by Van Der Schaaf, is ‘any situation in which an ongoing sequence of events was prevented from developing further and hence preventing the occurrence of potentially serious (safety-related) consequences' (as cited in Lucas and Hale 1991, p. 5). Near miss reporting can provide a range of insights that include identifying new hazards, statistically quantifying factors that give rise to accidents and gaining qualitative insight into how failures develop.

If examined collectively, incident data can also reveal recurring patterns of incident-producing circumstances and lead to a strengthening of defences over time (Grote 2009; Ale 2009). Aggregating data from across an industry allows for a greater opportunity to learn about even rare events. However, in more recent years the ongoing utility of collecting and reviewing fatality data for preventing future accidents has been questioned.

Maurino (2000), for example, expressed the concern that, although important, focusing on accident data may not be the best use of resources, arguing that: 'While accident investigation must be recognized for its historical contribution to aviation safety, the industry cannot afford to use up meagre resources in reactive endeavours' (p. 958). Reynard (1991) suggested that, despite the anticipated potential of incident reporting programs, they have failed to live up to their anticipated potential (p. 2). Bourrier (2002) concluded that expanding databases to collect increasing amounts of information about accidents, incidents, near misses, and even precursors of near misses, was resulting in huge amounts of data and effort for not much benefit (p. 4).

Dekker (2014) refers to this problem as Mission Creep in the reporting and documenting of incidents, which now extends to 'precursors of precursors' (p. 350). Amalberti (2001) argued that the trend to look ever further back in time resulted in the creation of too much irrelevant information, with overly onerous reporting resulting in 'bloated and costly reporting systems with not necessarily better predictability, but where everything can be found; this system is chronically diverted from its true calling of safety' (p. 113).

Furthermore, Dismukes (2010) suggests that the predictive power of these databases is limited since 'the accident that did happen may have been no more probable than many that did not happen' (p. 335). Dismukes appears to be suggesting that accidents are the result of a rare combination of unique events, no more likely to combine again than any other combination of rare events. However, an alternative view might be that the accident that did happen is now less probable in future, precisely because it was recorded in the accident database and any related lessons were already learned.

Indeed, several studies have suggested that many work-related fatalities were foreseeable from relatively few repeating patterns, revealed in previous incident data (Laugherty 1993; Feyer and Williamson 1991). For example, in a review of fatal accident data by Williamson and Feyer (1991), a striking characteristic of their findings was that there were relatively few patterns of causal pathways found in the data. Twelve patterns were found, and the majority of the accidents were accounted for by one of four major precursor event sequences (66% of all cases in the sample) (p. 310). This finding suggests that many accident sequences do indeed combine and repeat in similar ways to contribute to accident causality. Clearly, learning from previous events provides an important contribution to safety management today. Incident data will no doubt continue to be an important source of information to prevent future harm, however, field of safety science has long sought more proactive measures of safety, which did not rely on reacting to past events. The concept of a 'leading indicator' gained popularity as a more proactive way to measure safety.

4.1.2 Leading indicators of safety

Reason (1991) argued that while incident and accident reporting systems are a necessary part of any safety information system, they are by themselves insufficient to support effective safety management. The information they provide is both too little and too late for this longer-term purpose. In order to promote proactive accident prevention, rather than reactive 'local repairs', it is necessary to monitor an organisation's 'vital signs' on a regular basis (Reason 1991, p. 9, in Van der Schaaf, Lucas and Hale (eds) 1991). Reason argued that one reason for this is that the data lags behind the event in question. Flin, Mearns, Connor and Bryden (2000) report a 'movement away from safety measures purely based on retrospective data or lagging indicators' such as fatalities, lost time accident rates and incidents, towards so called leading indicators (p. 177) Several others, like Hale (2009), have made note of the distinction between leading and lagging indicators – lagging indicators are reactive, only able to suggest corrective action, after the event has occurred (Hale 2009, p. 1). By contrast, Hopkins (2009) explained, leading indicators are thought to 'provide feedback on performance before an accident or incident occurs' (p. 5).

The deviation models of the 1990s (Kjellen and Hoveden 1993; Stellman 1998) encouraged the recording of upstream measures of performance, such as the status of safety prevention measures or the number of safety audits conducted. From the late 20th Century forward, there has been much discussion about which data make good indicators or early warning about safety performance (Powell et al. 1997; Kjellen 2009; Hopkins 2009; Hale 2009; Fairfax 2012). Wreathall et al. (1990) explored how properly developed indicators might serve as signals to management, which then allow them to take appropriate control over a given situation or initiate corrective action. The problem, however, has been determining what signals to record and attend to.

In a review of safety indicator literature, Swuste et al. (2016) describe over 23 different categories of indicators to choose from, which include measuring unsafe behaviours, the number of safety observations recorded, frequency of safety meetings held, numbers of safety inspections conducted, or training courses completed (p. 167). Clearly, having too many indicators could mask the importance of any particular signal, and a great number of fluctuating measures can make it difficult to know which ones are a matter of concern or should be investigated. Oien, in several works published between 1996 and 2011, noted that indicators must also be sufficiently motivating so that those in a position of power will be moved to take necessary action. As Oien et al. (2011) have suggested, even when warnings are received in the form of indicators, such warnings are often not acted upon because a causal link often goes unestablished until after harm has occurred. Cook and Woods (2008) referred to this as the problem of developing foresight: ‘It is not that data about leading indicators is unavailable; rather, what will be seen as clear warnings after the fact were discounted before the fact’ (p. 1).

To illustrate this point, Hopkins (2007) cited database entries from a BP Texas fire that demonstrated that managers were aware of safety indicators, such as recorded spills, getting increasingly worse in the two years preceding the event. Although several managers had access to this data, none were sufficiently provoked to respond in a way that prevented the accident from occurring (p. 12). This suggests that additional research is required to determine how organisations use safety indicators, why staff or management might ignore warnings that appear to be crucial, and whether this demonstrates a limitation of the indicators themselves or hints at a particular safety department’s lack of process, motivation or inadequate safety culture to take the required action.

The problem is recognising from the indicators when normal performance deviates from what is acceptable. Based on the deviation models discussed in Chapter 3, Kjellen (1984, 2000) argued that if we understood how operations normally functioned safely, we may be better able to identify signs of deviation. Svedung and Rasmussen (1996) suggested that, as well as learning from incidents and safety indicators, complementary methods, which increase understanding of how normal behaviours drift toward the boundaries of safe or unsafe performance, may be necessary. Bourrier (2002) agrees that it is normal operations that require closer scrutiny.

In deciding what are the right leading indicators to measure safety, Swuste et al. (2016) conclude ‘it is clear that the “silver bullet” has not yet been identified’ (p. 162). Bourrier (2002) also describes the debate about what to measure as ‘far from being settled’ but suggests in the meantime, that developing our understanding about ‘normal operations’ and how every day challenges are met may contribute to its maturation (p. 173).

4.1.3 Safety measures based on the Reason model

Attempts have been made to use Reason’s model as a proactive process measurement, where the assessment of a limited set of ‘vital signs’ gives an indication of safety-health and suggests factors for remediation, such as through the TRIPOD method (Reason, Wagenaar, Hudson, Benson and Groeneweg 1988–1990). A more recent attempt to convert the Reason model into a data collection tool has been based on the Human Factors Analysis and Classification System (HFACS) (Wiegmann and Shappell 2003). The ATSB (2010) noted that the HFACS was originally developed as a hierarchical taxonomy to describe the human and other factors that contribute to an aviation accident or incident. It has also been adapted as an audit tool to predictively identify where an organisation might be exposed through, for example, latent failures or holes in organisational defences in the absence of an accident.

Hsiao et al. (2013), for example, attempted to use the HFACS model as an audit tool and then analyse whether the audit results were predictive of safety outcomes in aviation maintenance. Their preliminary study found some moderate correlations between audit findings based on HFACS categories and subsequent safety performance in maintenance incidents involving two airlines. They concluded that HFACS-based auditing showed some encouraging results but that more research was necessary.

In a later review, the ATSB tested the relationships between HFACS taxonomy factors, using an analysis of Australian aviation accidents to assess the usefulness of HFACS as a predictive tool. The ATSB later suggested that HFACS, used predictively, may help safety improvement, but concluded that 'HFACS may have limited effectiveness, as a predictive framework' (Inglis et al. 2010, p. 33).

Whilst the Reason model works well as a framework for accident investigation, it has been less successful when applied as a framework for proactive measurement (Shappell and Wiegmann 2001). The difficulty of applying such tools prospectively might be determining in advance which characteristics contribute most to safe outcomes. Reason (1990) himself conceded that latent failures may only become apparent when they combine with other factors to breach the system's defences, which makes them difficult to spot and correct in advance.

4.1.4 Measures of safety management system (SMS) performance

Since the advent of the safety management system, which was introduced in Chapter 3, organisations have sought to develop indicators to assess how well SMS elements are functioning. Scores of SMS elements can be combined to provide an indicator of overall SMS effectiveness. SMS indicators might include traditional reactive measures, such as incident data, as well as leading indicators from SMS audits. The UK Health and Safety Executive (HSE) (2001) recommends measuring the effectiveness of the SMS by how well it is able to identify hazards and the effectiveness of controls at managing risks. The HSE also recommends measuring the safety outcomes of the SMS in terms of incidents, injuries, accidents and near misses data (p. 22). Typical SMS effectiveness scores might also include scores of proactive safety related activities, such as the number of investigations completed, safety actions implemented, safety training programs attended and inspections and SMS audits conducted.

Attempts to combine such a wide range of data points into a single score, or index, have resulted in sometimes complex and cumbersome data collection tools. In 2006, the HSE proposed over 40 different indicators at three levels, to produce one overarching indicator of safety management performance. Bird and Germain (1985) proposed using more than 600 different data points to measure SMS effectiveness, which could then be summarised according to their relative weights to arrive at an overall score.

The proposed advantage of these combined SMS scores or indices is that they provide a baseline score of how an SMS is performing and that they continually generate data, which helps monitor SMS progress and improvements. Disadvantages include the large number of data points necessary to produce a baseline score and the difficulties inherent in identifying which aspects to focus on for improved safety performance. Kjellen (2009) argued such measures have resulted in too much detail and complexity, which make it difficult to identify the few vital indicators that should capture management attention. Index SMS scores are based on the assumption that if SMSs are in place and are continuously audited, the organisation will operate safely.

As a gross measures of overall system performance, SMS indices may provide a useful indicator of how well an organisation's safety management system has been designed and implemented. The amount of effort involved in setting up and monitoring each SMS element, to the extent necessary to produce an SMS score, is considerable and may not yield sufficient insight into the factors creating and detracting from safety in every day operations to warrant the required investment. Hopkins (2000) also cautions that whether the resulting SMS score is an accurate reflection of how safe an organisation is depends on whether an SMS is sufficient to assure safety in the first place. The concern is that detailed monitoring of existing safety practices tends to confirm the status quo, or even generate a false sense of security, rather than provide the feedback loops to drive innovative opportunities for improvement originally envisaged by SMS proponents.

Paris and Haywood (2016) argue that the vision of a safe organisation suggested by most SMS is one of ‘a fully controlled world in which managers and designers anticipate every potential situation, predetermine all the appropriate responses and full control over frontline operators – who (should) consistently follow the corresponding prescriptions for success’ (p. 211). They argue, however, that this view has been challenged by both past and contemporary research on organisational reliability, such as normal accident theory (Perrow 1999), the High Reliability Organisation (HRO) movement (La Porte 1996; Roberts 1990) and the Resilience Engineering concepts. Paries and Haywood argue that in most industrial processes, strict adherence to pre-established action guidelines aspired to within the SMS model is ‘unattainable, incompatible with the real efficiency targets and insufficient to control abnormal situations’. They describe this problem as ‘the irreducible difference between prescribed work and real work or work as imagined and work as done’ (Paries and Haywood 2016, p. 16)

4.1.5 Use of incident data, indicators and measures of safety systems in aviation

Aviation has a long and successful history of collecting data and making incremental safety improvements informed by incident-based insights (Mahajan 2010). The FAA’s newest safety reporting tool, the Aviation Safety Information Analysis and Sharing (ASIAS) program, allows users from across the industry to perform ‘integrated queries across multiple databases, in an extensive warehouse of safety data’ (FAA 2014, p. 1). Ground operations, in particular, also collect a large number of health or safety indicators and industry-specific indicators in relation to airport worker injuries, aircraft ground damage as well as unsafe ramp conditions.

Domain-specific databases are managed by regulating bodies such as the ICAO, IATA, GAP and Global Safety Information Centre (GSIC). However, despite the large amounts of data available in terms of safety performance, ground operations still lags behind the rest of the industry. Drury et al. (1998) analysed interventions developed in response to airline ground-incident data only to conclude that 'airlines have been generally ineffective in planning, developing and evaluating countermeasures to address recurring safety concerns in the ground operations environment based on incident data' (p. 89). So, whilst incident data is an important source of information, it has not yet informed the improvement necessary to keep pace with the rest of the aviation sector, suggesting that new, complementary, approaches might be necessary (Balk and Bossenbroek 2010; Passenier 2015).

The Reason model has been extremely influential in aviation, as discussed in Chapter 3, but mainly as an accident investigation tool. Measuring safety with tools based on the model can be problematic, as it is difficult to identify latent conditions in advance. Reason (1996) recognised some of the difficulties, describing the Swiss Cheese model as 'a weakly predictive model' which can be used as a form of measurement for assessing the general health of a system, but agreeing that attempts to use the model as a proactive measurement tool have 'met with limited success' (p. 20).

Safety management systems are now a requirement for most airlines, as discussed in Chapter 3. In 2008, the IATA began developing a ground safety SMS audit tool known as the International Safety Audit for Ground Operations (ISAGO), in an effort to provide a global standard for evaluating the SMS performance of ground handling organisation. Its stated aim is to 'improve operational safety in the ground operations environment, reduce damage to aircraft and equipment, and improve efficiency by reducing the number of redundant audits' (ISAGO 2010, p. 2). Although no data is yet available on the success of the ISAGO program in achieving its aims, it is likely that SMSs will continue to dominate the ground aviation safety management landscape for some time to come.

4.2 Self-Report Data

The information processing models and SHELL models discussed in Chapter 3 highlighted the need to consider the role of the human in creating and maintaining safety. Identifying human performance issues from incident data or leading indicators can be difficult. Few leading indicators have been developed for this purpose and extracting human performance data from lagging incident data is also problematic. Self-report measures refer to data provided by those individuals who are the subject of the research. They rely on the participant to report their own behaviours, attitudes or beliefs. Typical forms of self-report data in the safety domain include confidential incident reporting and surveys and interviews. Each is discussed below in relation to measuring safety.

4.2.1 Confidential reporting

Reynard (1991) noted that when interviewing participants in events they may be unwilling to outline their own actions and errors that contributed to the safety failure – particularly if doing so is personally disadvantageous or has the potential to elicit a penalty (i.e. discipline, blame, prosecution) (p. 11). Recognising that valuable information about incidents was being lost, several industries have set up confidential reporting systems to protect those filing submissions and encourage robust reporting of events, near misses and/or unsafe acts and conditions that operators identify as having the potential to cause harm (see Barach and Small (2000) for a review).

Confidential reporting systems allow participants of events to provide information under the protection of confidentiality. One of the longest running self-reporting programs was developed in aviation. The Aviation Safety Reporting System, or ASRS, is maintained by NASA to provide an archive of aviation safety incident reports that are voluntarily and anonymously submitted by pilots, air traffic controllers and other personnel who may have relevant information regarding safety incidents. O'Leary and Chappell (1996) and others have cited similar systems that were developed elsewhere, such as the Confidential Human Factors Incident Reporting (CHIRP) used in aviation and maritime settings and REPCON (Rail Voluntary Confidential Reporting Scheme), the ATSB's rail workers' voluntary, confidential reporting scheme used in Australia.

The main advantage of such schemes is to provide a mechanism for accessing data from the perspective of those involved, which is unlikely to be available through any other means. Another advantage of self-reporting is its ability to provide richer, qualitative data about a wider range of incidents and near misses than accident reports alone. Confidential self-reports have also generated event-specific information that was otherwise not reported through official channels. Reynard (1991) has noted that, in several cases, confidential reports were virtually the only source of incident data provided (p. 3).

Disadvantages of self-reporting include its dependence on voluntary submission, which may not be forthcoming, and subjectivity. In a study conducted by Wu et al. (2007), which evaluated healthcare professionals' intention to use adverse-event-reporting systems, several factors were found to affect a willingness to report, including for example the perceived usefulness of the information being submitted, the reporting systems' ease of use and beliefs about whether or not reporting would effect change. Each factor had a significant impact on intention to use reporting systems and the information provided (p. 1).

Self-report data is, by definition, information from the reporter's point of view. This may mean it is selective and biased in terms of what respondents are willing or able to report. Weick described the problems people experienced when trying to 'make sense' of uncertain situations or things that happen to them (Weick 1995, p. 9), which can lead to distortion when reporting subjective experiences. Selection of what gets reported may depend upon what the reporter considers to be important whilst missing other details that may appear to be insignificant. Errors that on one occasion resulted in few or no significant consequences may go entirely unreported, for example, even if the same error in different circumstances could have catastrophic consequences. Therefore, researchers cannot assume that the data is representative of events that happen during normal, everyday operations.

4.2.2 Surveys and interviews

Another method for collecting self-report data about human performance is through surveys and interviews (Seidman 2012). Interviews can be used on their own or in combination with surveys to derive further detail from survey data. Like confidential reporting, they provide a mechanism to gather self-report data that would not be available from other means. Interviews allow researchers to gather in-depth information about how humans have performed in the past or are likely to perform in the future. Surveys can include more participants, but interviews offer the advantage of being able to probe in greater detail. Weiss (1995) explains how interviews might be tailored around respondents' experiences. The nature of interview questions, and the ways in which they are asked, can influence results (Bradburn 1980). However, well-structured interview and response coding processes can be employed to make the approach more rigorous and reduce bias (Frick, Christopher and Kamphaus 2010). Whilst interviews are more time consuming and resource intensive than surveys, they often yield richer and more detailed data – albeit from a traditionally smaller sample size.

Interviews can also be used to gather information about normal operations before incidents or accidents occur. Hays (2009), for example, interviewed operational managers from three high-hazard organisations and asked them to describe situations in which they had to balance or trade off safety concerns with efficiency needs in the course of normal operations. Using interviews, Hays was able to gather information about these day-to-day qualitative decisions and trade-offs, which might have been difficult or impossible to collect through other methods. Surveys and interviews are, therefore, both best employed in situations where information about the subjective experience is sought. They also provide a mechanism for collecting information that can be put into context and exploring, from the operator's perspective, what normally occurs and what occurs when things go wrong in the work environment being explored.

Surveys are often used when a large population of general perspectives is sought. Surveys can be used to solicit feedback on past experiences and to investigate perceptions about where harm might arise in the future. Surveys can also be administered anonymously, allowing respondents to provide information that might otherwise be sensitive and difficult to obtain. Surveys can be relatively inexpensive to administer, reach a broad range of the target population and generate large amounts of data from those with first-hand experience of both their own performance and their organisation's safety practices. They can also provide a representative sample of the larger distribution of problems experienced, which can then be statistically analysed for trends. In this way, survey results can serve as a useful baseline measure and mechanism for measuring trends over time.

Like all forms of self-report data, interviews and surveys are necessarily subjective, which has both advantages and disadvantages; however, in many cases it is the only way such information can be accessed. Responses may demonstrate biases influenced by a range of factors, including the organisation's perceived motivations for administering the survey and respondents' motivations to answer honestly. With all surveys, the instrument itself can sometimes influence what data is collected. Sampling methods and sample size are also important factors that influence the validity of data (Kotrlik 2001). Even the way a question is asked, and the options available for responding, will influence how people are likely to respond (Wilson 2005; Dillman, Sinclair and Clark 1993).

As with all self-report data, respondents may also lack the insight to respond to questions about their own human performance limitations. Helmreich (2000), for example, found that pilots downplayed or disregarded the impact factors such as stress and fatigue had on their performance. Nederhof (1985) found that, even with anonymity, respondents were often uncomfortable giving answers that reflected unfavourably upon them. Goritz (2004) noted that motivation or incentives for completing questionnaires can also influence the validity of results.

Sometimes it is precisely the selective, subjective opinions of participants that are being sought in self-report data, in which case interviews and surveys offer particular advantages. Where the perceptions or attitudes of the participants are thought to influence safety, surveys are the method of choice for sampling perceptions in the target population.

Safety Culture surveys. The Safety Culture models discussed in Chapter 3 have led to the creation of a wide range of tools for measuring safety culture and climate. Much research has been undertaken to try to deconstruct the elements that define a safety culture/climate and develop measurement instruments (Fernandez-Muniz et al. 2007; Zohar et al. 1997; Flin et al. 2000; Mearns et al. 2003; Hudson 2003; Reason 1997). Zhang et al. (2002) have noted that there is still considerable disagreement among researchers as to how safety culture should be measured.

Williamson et al. (1997) examined seven reports measuring safety climate and concluded that there were at least eight factors to consider: four measuring attitudes and four perceptions. Flin et al. (2000) concluded, after a review of several instruments, that there is in fact limited evidence for or against a common set of core features (p. 20).

Despite the wide range of measuring instruments available, there is still uncertainty regarding the way organisational cultures influence safety outcomes. Identifying cultural problems is perhaps less challenging, however, than knowing what to do to effect change that would bring about safety improvements. Heese (2012) stated that organisations repeatedly find it difficult to derive transformative action plans from information provided by safety culture or climate measurements. She suggests that, compared to the number of theoretical models available, there is little practical guidance on how to transform safety culture from theory into practice.

The High-Reliability Organisation (HRO) surveys. Ciavarelli and Figlock (1997) developed a survey tool based upon key principles of the High Reliability Organisation model discussed in Chapter 3. The survey asked military personnel to rate how well they believed the U.S. Navy and Marine Corps compared when measured against those HRO characteristics. Figlock and Ciavarelli attempted to demonstrate the predictive validity of the tool by comparing survey results with safety outcomes. They measured the rate of accidents that occurred in a defined period following the survey. Figlock and Ciavarelli found high-scoring naval, marine and aviation organisations to have predictably higher safety records and lower accidents rates. Indeed, the lowest accident rates were associated with units that earned the highest HRO scores (Ciavarelli 1997, 2005; Ciavarelli and Figlock 1997).

Weick and Sutcliffe (2005) offer a series of 'audits' or rating scales that assess the extent to which an organisation is behaving like an HRO and give some general advice about how to improve (chapters 5 and 6), but do not give us much insight into precisely how these goals can be accomplished. HRO surveys do not measure reliability directly but, rather, staff perceptions about the presence or absence of HRO features. As with any survey tool based on a model or theory of safety, the validity of the tool itself depends in part on the validity of the underlying model. There is also the problem of converting survey results into intervention strategies. Despite excitement and productivity from early researchers, publications on HRO practice are largely silent with respect to evidence that highly reliable operations can be attained, and once attained, can be sustained and positively impact safety performance (HSE 2011). Newquist, Tolk, Cantu and Beruvides (2015), for example, comment on the limited success of HRO approaches in the healthcare industry and report that, 'While several HRO frameworks have emerged… transformation remains elusive, improvements remain modest and patient harm continues to be pervasive' (p. 900).

The HRO model and survey tools offer interesting insights into features that are common among high-performing organisations and demonstrate a correlation between organisational features and safety performance. However, more research may be necessary to establish a causal link between the features described and the safety outcomes they are thought to affect. Marx (2017) points out this issue when referring the disparity between the highly reliable flight safety record and their poor performance in work place injuries. Suggesting that if commercial airlines already knew how to create HROs they would have used the same approach to 'make aviation a safe place to work, get schedule reliability above 90% and lose a few less bags' (p. 2), Marx suggests that organisations can only achieve extraordinary results around those things they choose to value. He proposes instead to '[d]esign great systems around employees. Build the right culture. And build in the systems of surveillance that allow the organization to learn' (p. 2).

Resilience measures. If resilience creates safety, as the theory and model discussed in Chapter 3 suggests, then surveys that measure an organisation's ability to behave resiliently could help to identify strengths and areas for improvement to prevent the next accident from occurring. Attempts to develop the model into a data-collection tool have focused on asking participants about their organisation's ability to demonstrate features of resilience, such as detecting and making corrections when things go wrong, being able to cope with the unexpected and both adjusting to and recovering from negative events. Woods and Hollnagel (2006) propose a number of safety indicators that could be gathered to assess an organisation's resilient qualities. Hollnagel (2010) incorporated similar concepts of resilience into the Resilience Analyses Grid (RAG), a tool that attempts to measure resilience by focusing on asking participants about the organisations ability to respond, monitor, anticipate and learn from failures. The resilience ratings are provided based on answers to questions for each of the four areas. Results are depicted in a graph that represents what Hollnagel has called 'the system's potential for resilient performance' (2015, p. 1).

Proponents say resilience theory will provide both models for how accidents occur and, eventually, measures of how well organisations are performing (Hollnagel and Woods 2006; Dekker 2006; Amalberti 2001). Woods and Hollnagel (2006) suggested that sufficient progress had been made in the field of resilience, as an alternative safety management paradigm, to begin deploying that knowledge in the form of engineering management techniques (p. 11). However, this assertion may be premature, since few actual tools, techniques or interventions have been tried and none have been convincingly demonstrated as effective in the current literature. Many would regard Resilience Engineering as an idea that is not yet proven and is even further away from being operationalised into valid safety management methods.

The concept of resilience engineering may yet be too underspecified to be measured empirically. McLeod (2012) has suggested that more evidence is needed to thoroughly understand and identify which resilient behaviours influence safe outcomes before a tool can be designed to measure them. Oxstrand and Sylvander (2010) saw no real difference between measuring resilience and measuring aspects of safety culture and, therefore, suggested that there are no benefits to measuring resilience per se. McLeod supported this, saying that without a clear understanding of what resilience is or how it functions, it might be more beneficial to first just learn more about what people do successfully every day to manage threats and failure when going about their normal work (AAVPA 2012).

As well as measuring features of an organisation, Resilience theorists have also suggested the need to analyse the difference between work as imagined and work as performed. Hollnagel et al. (2011) suggest that '[g]etting smarter about predicting the next accident (if we see this as part of resilience engineering) is in part about finding out about this gap, this distance between the various images of work (e.g. official vs real) and carefully managing it in the service of better grasping of the actual nature of risks in operations' (p. 92). Shorrock (2016) suggests that organisations should pay attention to work-as-done by studying recurrent patterns, flows, trade-offs and compromises, between this and how work is prescribed. Shorrock suggests that by exploring the gaps and implications, it may be possible to build more resilient systems

4.2.3 Use of self-report data in aviation

Although aviation has also been world-leading in the use of confidential reporting and data collection tools, the databases contain very little information from ground handlers. NASA and the FAA implemented the ASRS program initially developed for flight crew; this was later made available to maintenance and flight ramp personnel. As with any form of self-reporting, however, these programs rely on participant's willingness to report (Steckle 2014). Fewer reports are received from ramp crew members than from members of any other aviation subgroup (Taylor 2013). Consequently, there is far less ground-incident data compared to other areas of aviation, such as flight safety and maintenance. It is not immediately clear why ground handlers do not make use of such channels, but it would appear the data available to date is insufficient and new approaches may be necessary to understand the human performance issues from the ground handler's perspective.

Aviation has a long history of using surveys and interviews to collect data about human performance, factors that influence pilot performance and perceptions of safety culture (Harris et al. 2005; Helmreich 1987; Sexton et al. 2000; Sutton 2012; Von Thaden et al. 2003; Patankar et al. 2003; Gibbens et al. 2007). Questionnaires and interviews have also been used, in ground and maintenance environments, to understand which errors are perceived to occur most often and which performance-shaping factors are perceived to be the most influential (McDonald 2000).

Balk and Bossenbroek (2010) have conducted surveys and interviews with ramp personnel. The objective of this study was to investigate the causal factors which lead to human errors during the ground handling process from the perspective of ramp workers and their managers. The study concludes by identifying several target areas for improvement based on their results, including safety policy and principles, just culture, communication management supervision and commitment, standardisation of phraseology, and awareness of the potential risks from threats such as time pressure, stress, fatigue and communication. This study offers interesting insights as to what might be contributing to safety issues on the ramp from the perspective of ground handlers. Future research could seek to identify more objectively the extent to which these factors may be influencing safety performance and where safety efforts would be best targeted to bring about improvements.

Several attempts have been made to measure and influence safety culture in aviation generally including ground-based areas of aviation such as maintenance and ground handling (McDonald, Corrigan and Cromie 2000; Ek and Akselsson 2007). In a study of nine dimensions of safety culture carried out at a Swedish airport, Ek and Akselsson (2007) found that average safety scores were generally lower for the ramp than for other, comparable, studies in the air traffic control sector (carried out by Ek et al. 2002) and the shipping industry (Ek and Akselsson 2005). Balk and Bossenbroek's (2010) survey of ramp workers discussed above also investigated perceptions of safety culture dimensions on the ramp and made several recommendations for improvement.

Influencing culture in ground handling operations, however, has been complicated by the fact that multiple, complex organisations operate on the ramp (Schwartz 1999). Researchers like Guldenmund (2010), Hale and Guldenmund (2010) and de Boer (2013) recognised this problem – with De Boer (2013) stating that traditional safety-enhancement tools, like safety culture models, are 'limited in their effectiveness within this challenging domain, safety culture surveys may identify areas of weakness but have limited diagnostic capability and the correlation between safety culture factors and safety performance still awaits a fundamental scientific underpinning' (p. 1).

Aviation is already considered by many to be a high-reliability industry (Amalberty 2001) and aviation and the industry has inspired early HRO studies such as Rochlin La Porte and Roberts's (1987) HRO research into naval aircraft carrier air operations. However, these studies were referring to flight aviation rather than ground safety, which lags behind the rest of the industry in terms of safety performance (Balk and Bossenbroek 2010; Passenier 2015). Ciavarelli (2013) reported that surveys to measure HRO features have been implemented in military aviation, which included military ground operations, although this data is not publicly available.

Resilience measurement approaches have also been attempted in aviation, such as use of the RAG in air traffic control (Ljungberg 2013). Using the RAG tool, the Air Navigation Service Provider in Sweden (LFV) attempted to measure its resilience through interviews of air traffic controllers. However, there are no published attempts to measure high reliability or resilience in ground safety. The concept, however, of trying to measure behaviours that contribute to safety as well as detract from it is perhaps an important consideration for future research.

4.3 Observations

4.3.1 Direct observation

Although some safety data is only accessible through self-report measures, observation can offer an alternative data collection tool, particularly when seeking to capture information about how tasks are performed in everyday operations. The influence of the information processing and SHELL models can be seen in the use of observation tools for collecting data about human performance and interactions with technology such as errors, equipment use or handling issues and issues with the design of the human/machine interface.

Observational methods were pioneered by psychologists studying human behaviour and interactions, wherein they used a narrative (or detailed, written descriptions) to record their observations of human activity. Analysis of observation narratives became a method of choice in early naturalistic research (Barker 1963; Baumrind 1967; Wright 1967; Parke 1979). Unlike self-report data, observation studies do not provide insight into why individuals perform in a certain way, but they can be used to document how humans behave and interact with their environment.

Observation studies have also been used to analyse human interactions with products and technology. Barber and Stanton (1996), for example, used observation and coding of human error to analyse human interactions with a new ticketing machine. Observational techniques have been utilised in a broad range of human factors and ergonomics studies, including those related to human performance and behaviour in aircraft cockpits, manufacturing, control rooms and medical environments (Degani and Wiener 1997; Cook and Woods 1996; Seagull and Sanderson 2001; Bisantz and Ockerman 2002; Gaudart 2000; Mumaw et al. 2000; Stanton and Ashleigh 2000).

In ergonomics and human factors studies, the method most commonly employed involves an observer who is visible but avoids interfering with normal behaviour (Lincoln and Guba 1985; Wilson 2004.) Most observational methods are based on a conceptual model to represent the human performance being observed and categorise the data. More structured observations may use checklists or rating schemes to record the occurrence of predetermined activities (Dalijono 2006; Frey 2009). Less structured observations are sometimes based on qualitative descriptions, wherein observers record activities they judge to be relevant

(Blehar, Lieberman and Ainsworth 1977; Bisantz and Drury 2005). In the narrative method, observers provide detailed and comprehensive descriptions of everything the subject does over a given period of time. Once a number of detailed observations have been conducted, the researcher may then code the data for the purpose of performing quantitative and/or qualitative analyses. Observations typically record other factors that can impact performance (i.e. time of day, environmental conditions, light, noise) and look for interactions with human performance.

Observation methodologies can be labour-intensive and time consuming (Wright 1967; Dalijono 2006). In addition, observers usually require extensive training and calibration against standard coding mechanisms (Bao 2009). Even with extensive training, ensuring reliability and consistency between observers can be problematic. Observers have to determine which aspects of human performance they pay attention to, which can introduce bias and subjectivity, as can deciding what information to analyse and how it will be interpreted (Wilson 2004; Huberman 1984).

Another serious concern is whether or not subjects will behave normally if they know they are being observed (Johnson and Bolstad 1973; Johnson 1975; Haynes and Horn 1982; Kazdin 1982). It is possible to observe subjects covertly, but it has been noted that covert studies are fraught with ethical issues and can only be used appropriately in specific settings, such as when assessing pedestrian behaviour (Wilson 2004; Cinnamon, Schuurman and Hameed 2011). In most observational studies, subjects are aware they are being monitored and the presence of observers may distort normal behaviour (Rosenberf 1969; Patterson and Sechrest 1982). Bisantz and Drury (2005) suggested a number of strategies for avoiding observer effect and reducing distortion levels, which include having the observer: explain their role upfront, remain in the setting for an extended period of time, and focus on the work itself rather than those performing the activity. Even so, Bisantz and Drury conceded that it is 'unrealistic to think that people, cognisant of the presence of the observer, may not change their behaviour' (p. 78). The question is, perhaps, not whether distortion can be avoided altogether but whether valuable data can be recorded in spite of these concerns.

One safety program that has dominated the aviation safety landscape is based on the assertion that observations can collect very valuable data about human and safety performance even when subjects are aware they are being watched. This observation data collection tool is based on the TEM model explored in Chapter 3 and is known as the Line Operation Safety Audit, or LOSA.

4.3.2 Use of observation in aviation - the Line Operation Safety Audit (LOSA)

LOSA is an observation methodology developed specifically to study the performance of pilots in a naturalistic setting during normal operations. Although developed to measure the performance of flight crew teams, LOSA also takes into account the influence and interaction of the broader operation and organisational system. LOSA observations are usually conducted by a mixture of outside and peer observers and are based on the naturalistic approach. Subjects are aware of the observer, who tries to be as unobtrusive as possible.

The LOSA method is based on the Threat and Error model of human performance introduced in Chapter 3 (Helmreich et al. 1999; Klinect et al. 1999). According to the TEM model, threats and errors can either go undetected, be effectively managed or result in additional errors that require subsequent detection and response. If crew response is ineffective, the consequence of threats and errors may result in undesirable aircraft conditions, which pose an increased risk to safety. In a LOSA audit, pilots who are trained as observers observe flight deck operations and provide feedback about a flight crew's threat and management skills during routine, scheduled flights.

Observers record a linear narrative of the crew's activity and later enter that narrative into a database for the coding of activity according to threats, errors, any UASs and how each was managed. The analysis can identify which errors occur more frequently than others, which locations or aircraft types emerge as more problematic than others, and particular procedures that are more regularly ignored or modified by crew members. Interventions can be designed to target problems identified and repeated monitoring can be used to measure the success of safety improvement interventions. LOSA has the imprimatur of the aviation industry's agency, ICAO, which has published an Advisory Circular on LOSA implementation (ICAO Doc 9803 2002); this recommends some specific processes are put in place to ensure that subjects can behave naturally under observation without fear of repercussions. It is recommended, for example, that a secure data repository be used to protect the identity of the crew, ensuring complete confidentiality and non-jeopardy conditions. LOSA observers must not record names, flight numbers, dates or other information that would make possible identification of a crew associated with an observation. Data consistency checks are also built into the audit process to ensure that codes are consistently applied. This process is called data washing or data verification, where a small team of observers review the raw data codes for inaccuracies. Where the review team disagrees with the original observers' codes, a process of consensus is used to achieve agreement. Once the data has been checked and verified, it is ready to be analysed for trends.

LOSA data is examined for the prevalence and management of different types of threats and errors. ICAO (2002) has suggested that LOSA data be cross-referenced with other data sources, such as electronic data acquisition systems (e.g., Flight Operations Quality Assurance FOQA) and voluntary reporting systems (e.g., Aviation Safety Reporting System (ASRS). When available, LOSA reports may also refer to general findings from surveys or interviews that may aid further investigation of the causes of issues identified in LOSA data.

Loukopoulos and Dismukes (2002) describe one of LOSA's advantages as its ability to sample a large representative cross-sample of the operation using multiple peer observers.

Proponents of the LOSA approach, such as the FAA, have described it as a tool for collecting data about the full range of activities that take place in normal operations, including both the successful and unsuccessful behaviours of humans and their interaction with the system (FAA 2006; Pettersen et al. 2008). Ma et al. (2011) point to the proactive nature of this approach in finding weaknesses in the operation before they can cause harm and encouraging continuous improvement, which aligns well with safety management system concepts.

The LOSA observation methodology has been hugely popular in aviation, amongst carriers and regulators alike (Eames-Brown 2007; Klinect 2005; ICAO 2002; FAA 2006). While it has been primarily employed in flight operations, it is also used – albeit to a lesser extent – in maintenance, and has been increasingly employed in ground operation environments (Ma et al. 2011; Ma et al. 2012). The ATA's Human Factors Task Force and FAA researchers have completed preliminary development of a maintenance and ramp version of LOSA, known as M-LOSA and R-LOSA respectively; the results of early trials of these tools are discussed in more detail in Chapter 5.

At first glance, LOSA appears to offer an ideal solution. It is a proactive data collection tool, which considers both the human interaction with the system, in normal operations, under naturalistic conditions. The data collected is objective, representative, and informative in terms of identifying targets for change. LOSA is not without its critics, however, with some questioning the validity of the underlying TEM model and coding taxonomy, the observation methodology and the conclusions that can be drawn from the data collected (Dekker and 2001; Dekker 2003).

4.4 Summary and implications for the current research

The discussion above highlights a large number of potential measures that could be collected to improve our understanding of safety in ground handling environments, each with its own advantages and disadvantages.

For example, it would seem that incident databases will continue to be an important source of information to ensure we learn from past events. However more proactive measures are sought to identify problems before they cause harm.

Self-report data offers new insights into human and safety performance from the perspective of operators. Confidential reporting has yet to provide sufficient data from the perspective of the ramp. Both surveys and interviews offer richer complementary sources from the operator's perspective, but less insight into system interaction issues. The Reason model has been useful for accident causation, but less successful as a proactive measurement tool. Measurements of culture, resilience and HRO may inspire standards for ground safety to strive toward, but are currently underspecified as valid data-collection tools. New approaches to SMS that attempt to create feedback loops to drive continuous improvement show promise. Ramp SMS audit tools, such as ISAGO, may in the future highlight where ramp safety interventions are needed, if the right indicators can be found.

Although no single indicator or measurement tool has emerged as ideal for ground safety, the discussion above has highlighted a number desirable features any data collection tool should have. Methods that can illuminate the causes of incidents from their precursors are important to prevent their reoccurrence, but where possible more proactive indicators are sought to try and identify problems before they can cause harm. Modern safety management systems require data that provides feedback to drive continuous improvement. Data collection methods that can take account of the operator's perspective, as well as their interactions with the system, would seem to be advantageous. Data collection tools should seek to provide information about how humans perform in normal operations, to understand how things can go wrong, but also how humans can demonstrate resilience to manage challenges to maintain safe operations most of the time.

Observations offer opportunity for capturing data about how humans perform in their natural environments, which is not available through other means (i.e. incident databases, confidential self-reports or surveys) and may lead to more objective data collection.

LOSA is already considered by the aviation industry to be a proactive observation data collection tool, which considers human interaction with the system, in normal operations, under naturalistic conditions. The aviation industry's own advisory and regulatory bodies, such as ICAO and the FAA, describe LOSA as a method that can provide a leading indicator of human and system performance with the feedback loops needed to drive continuous improvement (ICAO 2002; FAA 2006). Recent adaptations of the model for use in ground operations, such as the M-LOSA and R-LOSA, which were developed concurrently with this research, are encouraging and warrant further investigation (Ma et al. 2011).

Despite its popularity and apparent suitability, however, serious concerns remain regarding the validity of LOSA and TEM, and whether it measures what it purports to measure. Nevertheless, it is clear that many elements of the LOSA method could be useful to this study of ground operations, if its methodological concerns can be satisfactorily resolved. Chapter 5 will, therefore, focus on exploring the LOSA methodology in greater detail and identifying any methodological issues that would need to be overcome if LOSA is to prove a suitable data-collection tool with satisfactory applications in ground handling.

Chapter 5: Consideration of LOSA for Ground Operations

Chapters 3 and 4 explored safety measurement tools and identified LOSA as having a number of desirable features as a tool for use in ground safety. Chapter 5 offers a deeper review of the LOSA methodology including its early development and growth in aviation, its suggested benefits and applications, and lessons learned from adaptations of LOSA outside the cockpit, including a recent adaptation for use in ground operations. The aim is to identify desirable features of the LOSA method that could be developed for use in ground safety as well identifying areas where improvement is necessary.

5.1 The Development and Growth of TEM and LOSA in Flight Operations

5.1.1 A tool for assessing CRM

The development of LOSA began with the simple aim of being able to audit the effectiveness of Crew Resource Management (CRM) programs for pilots. CRM training was time consuming and expensive to provide, so organisations sought evaluation tools that would convince them of CMR's effectiveness and thereby justify the investment (Croft 2001). In addition, there was growing pressure for airlines to evaluate individual pilot performance against standards of competence in the areas of technical and non-technical skill. In 1991, the LOSA research project began and was funded by the U.S. FAA. By 1994, a partnership was established between the University of Texas (UT) Human Factors Research Team and Delta Airlines, with the goal of developing an evaluation tool for CRM (Klinect et al., 2003 Merrit and Klinect 2006).

What resulted was an observation method for collecting normative data on crew performance during regularly-scheduled flights. LOSA was thought of as a step forward in measuring training effectiveness 'on the line' or in during normal operations, rather than measuring attitudes or participant reactions to training in the classroom or simulator settings. The audits provided data about the strengths and weaknesses of crew members as measured against a predetermined set of CRM behaviours or Non-Technical Skills (NTS) (Helmreich et al 1995, Flin and Martin 2001, Flin et al 2003).

5.1.2 The development of TEM and error taxonomy for LOSA

The LOSA tool evolved over time to keep pace with the shifting focus of CRM/NTS training. CRM training became increasingly focused on threat and error management, where pilots were encouraged to identify, trap and manage errors before they could cause harm (Matthews, 2004; Helmreich, Merritt, and Wilhelm, 1999). In 1997, the UT team – in collaboration with Continental Airlines – incorporated the coding of threats, errors and their management into LOSA observations. The mid to late 90s marked what Klinect et al. (2003) described as a paradigm shift in the approach to include threat and error management. Klinect, Wilhelm and Helmreich (1999) believed that 'Data are most valuable when they fit within a theoretical or conceptual framework' (p. 678), so they developed an early conceptual model of Threat and Error Management (TEM)

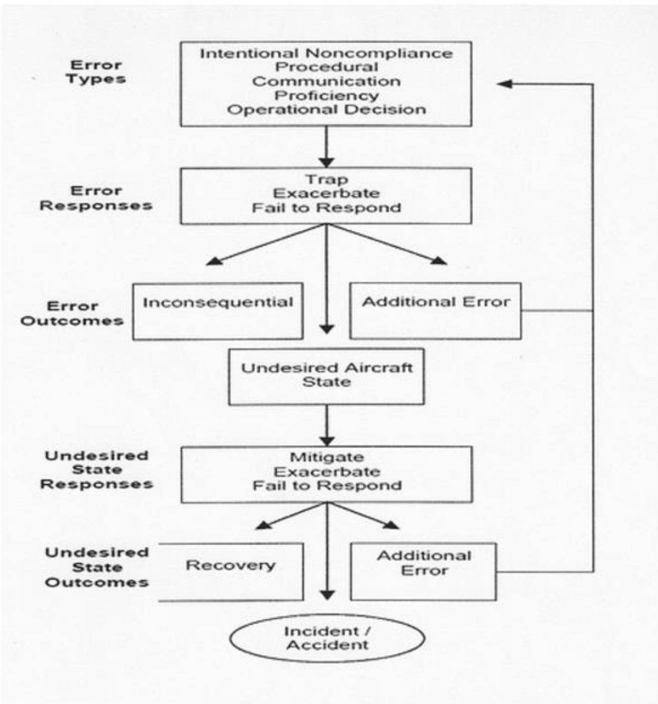


Figure 12: TEM Model as first depicted by Helmreich et al. 1999 (p. 3).

The model was used to describe and categorise the large amounts of data they were collecting from LOSA observations (Klinect, Wilhelm, and Helmreich 1999). Initial data suggested that threats and errors were common, but crew behaviours were critical to their effective management.

The terminology in the TEM model, described by Helmreich et al. (1999), is summarised with examples below. LOSA uses a five-category taxonomy to classify the types of errors that might occur, with examples from the flight deck context below.

- **Procedural errors:** when the intention is correct but the execution is flawed. It is meant to describe situations in which procedures were followed but incorrectly executed, such as when an incorrect altitude setting is entered.
- **Communication errors:** resulting from information being incorrectly transmitted or interpreted. Examples include incorrect read backs to Air Traffic Control, or supplying another pilot with erroneous course information.
- **Proficiency errors:** these arise from a pilot's lack of skill or knowledge, such as a deficiency in stick and rudder skills or an inability to program Flight Management System properly.
- **Operational decision errors:** these are discretionary decisions that increase the level of risk, such as unnecessary navigation through adverse weather patterns.
- **Intentional non-compliance:** conscious violations of Standard Operating Procedures (SOPs) or regulations. Examples include omitting required briefings or checklists, such as when performing checks from memory.

Helmreich et al. (1999) noted that there were three possible responses to the errors; they could be 'Trapped' – where the error is detected and managed before it becomes consequential; 'Exacerbated' – the error is detected but the crew's action or inaction leads to a negative outcome; or the crew could 'Fail to respond' – where the crew fails to react to the error either because it is undetected or ignored (p. 3).

Later, Helmreich et al. (2001) noted that after an error occurs there are three possible error management responses: *managed, mismanaged*, or *undetected/ignored*: Managed **s**ituations are those in which a threat or error is detected, trapped and dealt with successfully. Mismanaged situations occur when the threat or error is detected but the crew's response leads to a negative outcome that may then exacerbate the problem. Undetected/ignored situations occur when the crew fails to notice, react or respond to the error altogether. Threats and errors that are managed are described as *inconsequential*, in that there are no adverse effects on the safe completion of the flight. However, outcomes are said to be *consequential* if the crew action results in further error or undesirable aircraft states. The team at the University of Texas continued to refine the TEM model and its definitions and trial the methodology in live flight crew environments. A more detailed description of the development of the TEM model, definitions and LOSA methodology is provided in Appendix 2.

By 2002, International Civil Aviation Organization (ICAO) made LOSA a central focus of its Flight Safety and Human Factors Program and endorsed it as an industry best practice for normal operations monitoring (ICAO LOSA Manual, Doc 9803). ICAO recognises LOSA as a Standard and Recommended Practice (SARP) with the publication of Line Operations Safety Audit guidelines, which includes the following model of TEM in Appendix 2.2 of the document.

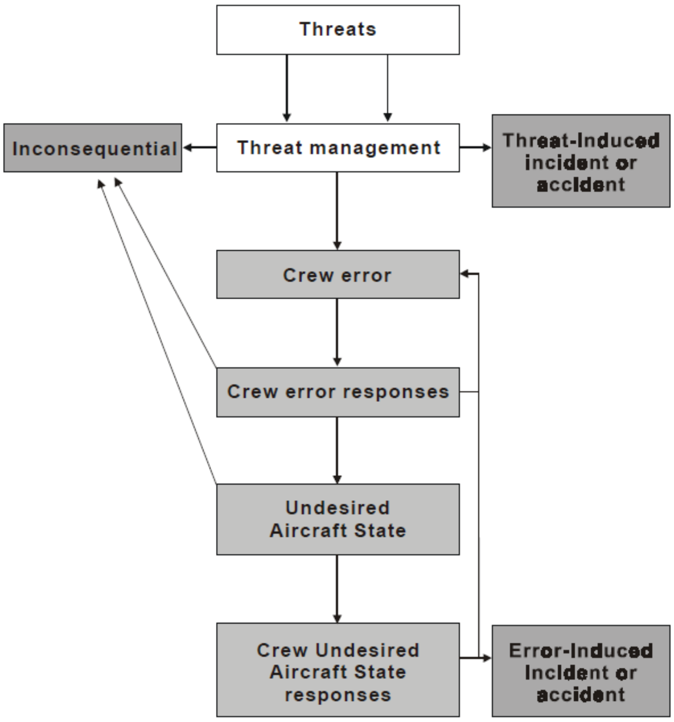


Figure 13. TEM model from ICAO 2002

This version of the TEM model and its definitions has become the most recognisable and the most commonly referred to by other subsequent descriptions of LOSA (see Appendix 2). The 2002 ICAO version of TEM included the five definitions of error as follows:

1. **Intentional non-compliance error:** Wilful deviation from regulations and/or operator procedures;
2. **Procedural error:** Deviation in the execution of regulations and/or operator procedures. The intention is correct but the execution is flawed. This category also includes errors where a crew forgot to do something;
3. **Communication error:** Miscommunication, misinterpretation, or failure to communicate pertinent information among the flight crew or between the flight crew and an external agent (for example, ATC or ground operations personnel);
4. **Proficiency error:** Lack of knowledge or psychomotor ('stick and rudder') skills;

5. **Operational decision error:** Decision-making error that is not standardised by regulations or operator.

The document also described the relationships between threats and errors in the model where '[t]hreats can link to errors and error can link to other errors or undesired aircraft states'. It described how errors that are caught in time do not produce negative consequences, whereas errors that were not caught and managed could lead to consequences such as other errors or undesired aircraft states. (ICAO 2002, Chapter 1, pp. 1-5)

It is interesting to note that in later publications, such as Klinect's 2006 doctoral thesis on the LOSA methodology, an adapted version of the model is presented.

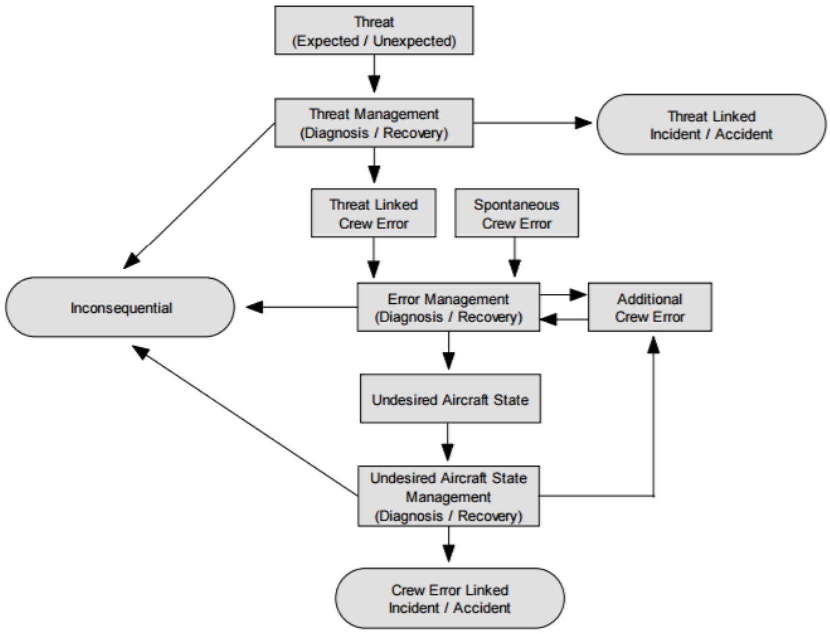


Figure 14: TEM model (from Klinect's doctoral thesis 2016, p. 16).

Klinect (2006) downplays the importance of quantifying the prevalence of threats, errors and UOSs and stresses that it is the management of these phenomena that should be of particular interest (p. 78). Klinect also notes that less than 10 per cent of cockpit crew errors had a threat identified as a precedent. Klinect also suggests that 'errors can be spontaneous, linked to threats or part of an error chain' (p. 17). There are also subtle differences in the threat and error definitions in the thesis, which are explored further in Appendix 2.

Such distinctions are of interest and are discussed further in section 7.1.2; however, the most influential publications of the LOSA method that followed, such as the FAA's Advisory circular 120.90 on 'Line Operations Safety Audits', refer exclusively to the ICAO 2002 TEM mode depiction. The FAA document does not reprint the model but references ICAO Document 9803 2002 as its source (further discussion of the examples provided in the FAA guidance document is provided in Appendix 2). Although there are subtle distinctions in the definitions provided, much of the reference documentation published by regulatory bodies refers to the original 2002 version. For example, the Australian regulator's advisory publication (CAAP 5.59-1(0) 2008) suggests that, as well as arising from external circumstances, threats could also arise internally (such as pilot fatigue), but the document nevertheless refers to ICAO 2002 Doc 9803 as its source documentation. The model and its definitions have been further interpreted and modified as LOSA tools have been adapted for domains beyond the cockpit, as discussed in section 5.3.

5.1.3 The growth of LOSA programs

Even before the endorsement of LOSA by industry bodies, the LOSA methodology was gaining widespread interest and acceptance in the aviation industry. UT established a LOSA Collaborative in an effort to provide support to airlines wanting to implement the program and compare their data against that of other airlines. By the year 2000, the BMJ (2000) reported that LOSA data had been collected on more than 3500 domestic and international airline flights. By 2006, the LOSA Collaborative reported that more than 30 commercial airlines in more than 14 countries around the world were applying the LOSA concept (Merritt and Klinect 2006). In 2016, the LOSA collaborative lists on its website over 45 airlines that have applied LOSA and further endorsements from the Flight Safety Foundation, International Air Transport Association (IATA) and the International Federation of Air Line Pilots Associations (IFALPA)

Among the reasons cited for LOSA's apparent success were the multiple applications for both the TEM model and the LOSA audit data. As well as providing data about normal operations, the TEM model was applied as a conceptual framework for training, a training evaluation tool, a model for incident investigation and an SMS audit tool (ICAO 2002; FAA 2006; BMJ 2000). The next section examines the merits of these applications in greater detail – with consideration given to whether such perceived benefits might be transferable to ground safety, if a similar LOSA-style approach were adopted.

5.2 Suggested Benefits and Applications of LOSA Programs

LOSA has been attributed several uses and benefits from its application in flight operations. The available evidence to support these claims is discussed below, as is the potential for providing similar benefits in ground safety if the tool could be adapted for this purpose.

5.2.1 LOSA as a leading indicator of safety performance

The FAA (2006) suggest that even a one-off LOSA study could provide a range of different indicators and measures with which to improve safety. They list nine specific indicators LOSA can provide that are identified as important to managing safety, such as the ability to:

- Identify threats in an airline's operating environment
- Identify threats from within the airline's operations
- Assess the degree of transference of training to the line
- Check the quality and usability of procedures
- Identify design problems in the human/machine interface
- Understand pilots' shortcuts and workarounds
- Assess safety margins
- Provide baselines for organisational change and
- Provide a rationale for the allocation of resources.

Much of the perceived usefulness, or face validity, of LOSA in the aviation industry has stemmed from the consensus that LOSA can provide the indicators and insights in these nine areas (FAA 2006; ICAO 2002), even though much of the evidence is anecdotal. There are still no published, controlled studies that demonstrate these particular benefits. It is likely that such benefits, if proven, would also be welcome in a tool developed to improve ground safety. The credibility LOSA has in aviation generally could also be a benefit for the acceptability of any new tools in ground-handling industry.

5.2.2 Use of LOSA data procedures and equipment design

Information about how crews interact with equipment in LOSA observations has anecdotally provided useful feedback for aircraft designers for the design of equipment and procedures. Boeing, for example, has claimed that LOSA data provides unique information relevant to the design of procedures and equipment (Applegate and Graeber 2005). Tesmer (2002) and Greaber (2006) describe how LOSA insights related to flight crew successes and errors have helped Boeing to create more user-centred and error-tolerant aircraft designs. Graeber (2006) described how LOSA data was used to confirm problems with procedures and checklists and improve their usability. Boorman (2001) and Holder (2004) also outlined how LOSA data has provided additional data and context to improve checklist designs

It has not been independently established that equipment and procedures based on LOSA data produce better usability or result in fewer errors than those developed without LOSA data or through other evaluation mechanisms. It is possible to say, however, that, using observational data, designers are able to learn more about how users interact with technology and procedures under realistic normal operating conditions. When considering potential benefits for ground safety, this type of information could inform the improved design of servicing equipment and SOPs; however, further research is needed to establish the value of data in this regard.

5.2.3 Use of LOSA to improve CRM behaviours

The CRM behaviours assessed by LOSA are based on previously established behaviours known as non-technical skills, or NOTECHS (Flin et al. 1998). Thomas, Sexton and Helmreich (2004) explain that while the behavioural markers monitored in LOSA are 'associated with aircrew performance [they] have not been studied employing experimental study designs' (p. 57). In other words, it has not been demonstrated empirically whether these particular CRM behaviours lead to safer flights. In a meta-review of CRM programs and their effectiveness by O'Connor et al. (2008), it was concluded that there is still insufficient evidence that CRM training achieves significant behaviour change or improves safety. The effectiveness of CRM programs based on the LOSA for improving safety outcomes is still unknown.

Croft (2001) argued that LOSA helps identify areas for training, thereby contributing to continuous improvement of training and the resulting human performance. Inherent in this cycle of improvement is the assumption that LOSA improves CRM skills that, in turn, improve safety outcomes (Thomas 2003, 2004; Lauber 1987; Wiener, Kanki and Helmreich 1993; Helmreich et al. 2001). However, this link has not been satisfactorily demonstrated (Salas et al. 2006). In an extensive review of CRM's effectiveness, Salas et al. (2006) were only able to conclude that CRM effectiveness at changing behaviour had 'mixed results' but they were unable to ascertain whether it had any impact on safety (p. 392).

Despite the lack of evidence, regulators such as CASA in Australia now require NTS training to be based on continuous-evaluation programs like LOSA (CASA CAO 82.5) The same behaviours have been adopted in training programs in other areas, such as air traffic control and healthcare, under the assumption that teaching the same topics will also be beneficial here (Sevdalis et al. 2008). Caution should be exercised, therefore, when considering the adoption of LOSA and its associated CRM behaviours for use in ground handling operations.

There is no evidence to suggest teaching the skills originally proposed for pilots would, in fact, result in safer ground handling crews.

Ideally, a new method is needed that is able to identify, based on evidence, which behaviours are successful or unsuccessful at managing risk in ground operations. Behaviours associated with positive safety outcomes could then be incorporated as targets for future training.

5.2.4 Use of LOSA in accident investigation

As well as its use as a predictive tool, TEM has been applied retrospectively as a tool for accident investigations. Helmreich (2006) has cited the use of LOSA data in several high-profile investigations. The TEM model had been used to describe how and where a threat or error was mismanaged during an incident, thereby helping identify recovery strategies and recommendations for improvement. Yet, the effectiveness of solutions identified through this process, in preventing reoccurrence, is unknown. ICAO (2002) also argued that LOSA could facilitate data sharing and learning from events, with the IATA Safety Committee adopting the TEM model as an analysis framework for its incident review meetings (p. 4).

Henry (2006) and Maurino (2008) claimed that LOSA data was being used as a complementary data source, in incident investigation, to identify how common it was for certain behaviours to occur. By investigating how commonly the error might occur in normal operations (even when no consequence results), LOSA data may help indicate the general scale of the problem, as well as help to identify better management responses. ICAO has recommended that LOSA data should be systematically used in this way, proving comparisons of normalized data to compliment accident-data trends.

Another perceived advantage may be the use of common terms for describing and classifying events for comparison. The use of a standardised model of human performance would be helpful for classification of actions in incident investigations, as well as facilitating the sharing and comparing incident data. An example of this is provided in a recent IATA report to support 'evidence-based training' or EBT (IATA 2014). LOSA data is used to complement other, more traditional, forms of accident data to understand the prevalence of 'threats, errors and undesired aircraft states encountered in modern airline flight operations' (p. 1).

When considering potential benefits to the ground handling industry, a model of shared terms could be useful for standardising data collection terms and improving shared learning from incidents. However, more research is necessary to establish the validity of the TEM model itself and to determine the value of TEM data to incident and accident investigations.

5.2.5 LOSA data as a leading safety indicator to improve safety performance

LOSA data is thought to provide a leading indicator of safety performance. As a proactive audit, and applied to normal operations, the data was thought to provide a leading indicator for safety performance. Helmreich (2006) called LOSA a proactive measure or early warning system. ICAO has said that LOSA programs enable operators to 'assess their level of resilience to systemic threats, operational risks and front-line personnel errors, thus providing a principled, data-driven approach to prioritize and implement actions to enhance safety' (2002, p. 2, section 2.1.2).

Both its creators and regulators have endorsed LOSA as a proactive approach that can prevent adverse events (Klinect 2006; ICAO 2002; FAA 2006). However, a link between implementing LOSA and reduced accident rates has not been independently established, although several airlines that have implemented LOSA programs have reported that LOSA has reduced incidents. Cathay Pacific, for example, has reported that LOSA data led to procedural changes that lead to improvements in stable approaches, flap-and-gear selection, take-off and approach briefings and landing callouts (as cited in Craig 2006). Tullo (2002) reported that Continental Airlines had experienced several safety improvements resulting from their LOSA program, including improved SOPs based on their LOSA results. Continental conducted internal follow-up studies after a LOSA program was introduced, finding a 70 per cent reduction in non-conforming approaches, although those numbers were not supported by published independent data (reported by Gunthur, cited in Klinect 2005 and Croft 2001).

Qantas (2008) similarly reported that its LOSA program helped highlight issues and hazards no other audit or monitoring systems had previously recognised, allowing them to be corrected. Qantas claimed, for example, that solutions based on their LOSA results have focused on reducing threats, improving problematic procedures and adjusting processes to avoid error-producing conditions. When the same issues failed to appear in subsequent LOSA audits, the airline's management team took that as evidence that the issues had been resolved.

Veilette (2008) reports that LOSA data has helped to identify problematic procedures and policies by highlighting poor adherence rates, which then became targets for improvement. Veilette attributes the flight deck LOSA program to numerous improvements in airlines, including the modification of dispatch paperwork, reallocation of resources and revisions made to procedures based on observation data.

Several airlines have claimed LOSA has provided unique and valuable data that has helped them reduce accidents and improve safety outcomes – although little published data is available to back up those claims (Klinect et al. 2003; Veilette 2008; Gunther 2002). So, despite the number of positive reports provided from the airlines themselves, it has not been demonstrated that airlines that have implemented LOSA have enjoyed better safety records than those that have not.

Nevertheless, the anecdotes and accolades indicate that LOSA enjoys considerable face validity and perceived value among industry leaders and regulators alike. LOSA's credibility in aviation could be an advantage for the introduction and acceptance of a new tool for ground the ground handling sector. However, claims that implementing LOSA reduces accidents require further research and validation.

Concerns about validity, however, have not hindered the growth and popularity of LOSA in aviation and beyond. ICAO has stated: 'Although initially developed for the flight deck sector, there is no reason why the [LOSA] methodology could not be applied to other aviation operational sectors, including air traffic control, maintenance, cabin crew and dispatch' (2002, p. 6). Since its introduction, LOSA has been adapted for use in all those environments. The next section will review the adaptations of the LOSA approach outside the cockpit and the lessons from these studies that are relevant to the current research.

5.3 Lessons Learned from LOSA Adaptions outside the Cockpit

The ability to adapt the LOSA model to other environments is an important consideration for the current research. This section reviews these adaptations to understand how transferable the method might be and identify any important lessons for consideration when adapting LOSA tools for the airport ramp.

5.3.1 LOSA adapted for dispatch and air traffic control

Henry (2006) adapted a LOSA-like approach to gathering safety data in dispatch operations. The approach remains faithful to the original LOSA methodology based on ICAO and Klinect (2005) in most respects (Henry 2006, p. 22); however, Henry (2006) provides further adaptation of the TEM Model, further details of which are provided in Appendix 2. Error types were adapted to the ATC environment to include, for example, position change errors, communication errors, equipment automation errors, flight data progress strips errors, procedural errors and aircraft instruction errors.

The related study involved training dispatchers to act as observers, who conducted a small number of observations and then coded both threats and errors and how they were managed. The study was supplemented by interviews and surveys. One objective of the research was to provide a detailed understanding of the ‘threat profile’ experienced by dispatchers; Henry reported, however, that difficulties were experienced in recording errors, as the task itself was not formally proceduralised and some errors were difficult to observe (p. 10). This study was one of the first to suggest that LOSA observation approaches might not be transferable to highly cognitive tasks.

This view is further supported by a study conducted by Dedale and Eurocontrol (as cited in Hollnagel and Amalberti 2001). Although not technically a LOSA, the method involved conducting observations and counting errors and therefore has some parallels with the LOSA approach in this study a model of human performance was developed, together with taxonomies needed to classify errors and contextual factors relating to Air Traffic Management incidents. The observers included an experienced air traffic controller and a human factors specialist/psychologist. Both were given three days training in applying the model and the taxonomy for coding observations. Following the observations, the data revealed that there were substantial differences in the number of errors noted by the two observers and – more importantly – only a very small number of errors noted by both.

This study demonstrates some of the inherent difficulties in studying highly cognitive tasks, but it also highlights another factor: the importance of context when observing behavior. Hollnagel and Amalberti (2001) suggested the reason for the disparity between observers was that the psychologists lacked the contextual understanding needed to interpret behaviours from the controller’s point of view, suggesting that observers require a detailed contextual understanding of the domain they are observing.

In a later adaptation of LOSA for air traffic control by Henry (2007), in which air traffic controllers were trained in the TEM model and then conducted observations over the shoulder of other controllers, context for interpreting results appears to have been less of a problem. Henry reported that the resulting snapshot of the operation provided rich contextual information and similar benefits to those claimed for cockpit LOSA as discussed above. Henry described how data was used in conjunction with other data-monitoring programs to support incident and accident investigations, inform the design of airspace technology and improve SOPs.

Applications of LOSA style observations in air traffic control highlight several important considerations when considering adapting the method for ground operations. Firstly, the method may not be suitable for highly cognitive tasks where actions and interactions are not observable. Ground crew activity involves observable actions, but communications between the team or with other departments, such as load control, would be difficult to overhear on the ramp due to background noise. The earlier study reviewed by Hollnagel and Amalberti (2001) has also highlighted the importance of context and observer experience when interpreting the behaviour and assigning codes. This would suggest that selecting observers from the target population is desirable.

5.3.2 LOSA adapted for rail: Confidential Observation of Rail Safety (CORS)

McDonald, Garrigan and Kanse (2006) adapted LOSA for train cabs and named the program 'Confidential Observations of Rail Safety', or CORS. The CORS approach refers to the original TEM model developed in Helmreich 1999 and outlined in ICAO 2002 (McDonald et al. 2006, p. 1) The authors describe how LOSA's ten operating characteristics were maintained for the study and how 'CORS observation materials follow a similar format to that of their LOSA equivalents however, the content of threat and error code lists is specific to the types of threats and errors that could be encountered in the rail domain' (p. 2). However, CORS did not include the observation of CRM behaviours (non-technical skills).

One objective of the study was to better understand TEM among train drivers so as to inform the development of CRM-style training for rail crews. The researchers wanted to use the data to 'learn which behaviours were beneficial' (p. 2) rather than attempting to measure CRM behaviours developed for pilots.

Train drivers trained as observers were positioned in train drivers' cabs to record threat and error management behaviours. A threat and error code taxonomy was developed, informed by a task analysis review of accident and incident data and focus groups, to identify the types of threats and errors that could be encountered in the rail domain. McDonald et al. concluded that the results of CORS were useful for highlighting future directions in training and for making improvements to organisational systems and processes (p. 4).

This adaptation of LOSA for rail highlighted possible approaches that could be used to developing new domain-specific taxonomies for ground operations, such as the use of task analysis and focus groups with subject matter experts to develop codes. The use of peer observers, and the preservation of the ten LOSA operating characteristics, may also be important for building confidence in the program and ensuring subjects can behave naturally. This study also reinforces the importance of extracting the ideal behaviours from the observation data rather than adopting untested behaviours from the flight operations CRM and NTS experience.

5.3.3 LOSA adapted for healthcare

LOSA has also been adapted into several healthcare settings. In 1993, Sanders and McCormick described one of the main benefits of the LOSA methodology in healthcare as being the provision of a standardised taxonomy, a data-collection method and a database. The applications in healthcare, however, also highlight a number of important considerations for adapting the tool to new domains, such as the need for extensive training for observers.

Jepsen et al. (2015), in a review of 23 behavioural-rating instruments adapted for healthcare settings, emphasised a need for increased knowledge about validating instruments, training observers and continuously refining those instruments. Jepson also suggested that further studies were needed to increase the validity of observation and coding instruments and to statistically investigate relationships proposed in underlying models.

When considering the lessons for the current research, the studies in healthcare suggest a need to consider how the training and tools can improve inter-rater reliability and to validate relationships proposed within their underlying models.

5.3.4 LOSA adapted for maintenance and ramp environments (M-LOSA/R-LOSA)

Several ground-based applications of LOSA have emerged concurrently over the period of this research project and are discussed in detail below. The first attempt at ground-based LOSA programs was initiated by Continental Airlines in 2008. Encouraged by initial results achieved at Continental, and other airlines, the FAA went on to sponsor additional research into the adaptation of LOSA to both maintenance and ramp environments. This led to the creation of the Airlines for America (A4A) Human Factors Task Force, formed with the aim of developing tools and a method for ramp and maintenance LOSA (Ma and Wood 2012). The resulting tools are known as R-LOSA and M-LOSA, respectively (Hackworth 2010; Ma and Wood 2012; Ma and Rankin 2012).

R-LOSA and M-LOSA are based on the TEM model and LOSA method and involve observations made during normal maintenance or ramp operations. Data is collected anonymously and confidentially by trusted and trained observers selected from the existing maintenance or ramp staff. Ma and Rankin (2012) propose that a Ramp LOSA program offers similar benefits to the nine first proposed by the FAA in 2006, outlined above.

The threats and errors are conceptually the same as in original cockpit LOSA, but adapted to describe the ramp working environment. Errors are defined as a ‘mistake that is made when a threat is mismanaged’ (FAA R-LOSA training, p. 6), but it is also recognised that ‘the threat-error linkage is not necessarily straightforward, and it may not always be possible to establish a one-to-one mapping between threats and errors. Errors can be spontaneous without direct linkage to threats' (Ma and Rankin 2012, p. 1). The methodology appears to remain faithful to the original LOSA /TEM approach in many aspects. Error definitions in the training program include the LOSA error definitions (intentional non-compliance errors, procedural errors, communication errors, proficiency errors and operational decision errors) (FAA R-LOSA training, p. 6). However, specific codes within these categories are defined as a deviation from a standard operating procedures, and therefore need to be customised for each environment.

Ma and Rankin recognise, that 'it may not always be possible to observe the threats that lead to an error' (2012, p. 61). Observers are asked not to speculate or guess what threats may have contributed to an error (p. 31); however, observers are asked to link threats to their associated errors outcomes (i.e. Inconsequential, Undesired State, or Additional Error). Observers also code errors as being Inconsequential, leading to Undesired State, or Additional Error (Ma and Rankin 2012, p. 40).

The process of collecting data differs from the original methodology. Flight LOSA relies on trained pilots using open-ended narrative text to record observations with much contextual detail. R-LOSA and M-LOSA utilise structured observation checklists filled in by observers with supporting comments. The use of checklists is favoured for the standardisation of data and for ease of use and is considered to be more effective than narratives in the ground domain (Ma et al. 2011, p. 7). Results are entered into a ready-to-use database and tabulated by data-analysis software.

The developers of M- and R-LOSA reported that the data verification process was retained, although somewhat altered, from the consensus review used in LOSA verification teams (ICAO 2002). In R-LOSA, data verification is undertaken by an analyst who reviews the codes assigned and the associated comments where available, rather than original narrative

(Ma and Rankin 2012, p. 16). The reporting of targets for improvement work in much the same way as they do in the original flight deck LOSA, with periodic LOSA monitoring used to determine if the interventions brought about the desired change.

The A4A Human Factors Task Force supported the development and testing of R- and M-LOSA at five U.S. airports in 2011, including a maintenance and ramp trial after which the task force released the R- and M-LOSA observation forms, procedures, databases and training materials. A number of successes were reported from the trial. The authors stated, for example, that M-LOSA observations were 'found to help make deactivation procedures more workable, efficient, and safer' (Ma and Rankin 2012, p. 41). Testimonials from ground handlers included in the trial suggest that the data helped to assess training programs and identify proactive trends that needed to be addressed. Participants from Airport Terminal Services (ATS) suggest that R-LOSA 'offers an excellent separate stream of data that ATS can compare to its audit trends. LOSA is a key link in the favourable safety results at ATS' (p. 51).

Rankin and Carlyon (2012) also reported anecdotally that the program helped make procedures more workable and avoided problems caused by shift changes. In addition, they argued that organisations involved in the study were 'reported to have reduced threats tremendously, as well as incidents of aircraft damage' (p. 12).

Ma and Pedigo (2011) concluded from their review that 'R -LOSA provides a minimally invasive safety audit of maintenance and ramp operations to evaluate an organisation including its systems, processes and personnel, to ascertain the validity and reliability of its information and consequently assess its internal controls' (p. 1). Rankin (2012) reported that R-LOSA provided 'a positive means for identifying threats and managing them before safety margins are reduced below acceptable levels. The result is a safer operation' (p. 12). Ma (2016) also provides further reports of positive Ramp LOSA implementations at United Airlines in the U.S., where R-LOSA data has been used to target improvements in employee/passenger injury rates and preventing aircraft damage, as well as improving compliance in areas such as GSE seatbelt use and belt loader operations.

In a recent application of R-LOSA, a research group for the Amsterdam University of Applied Science published details of an R-LOSA study conducted at Schiphol airport. In partnership with KLM, the research involved conducting more than 300 observations in ramp operations. De Boer (2013) noted that the tools required considerable customisation to suit the KLM/Schiphol ramp environment. For example, 25 error codes not supported by company or airport procedures were deleted and 12 new codes were added in. De Boer et al. (2013) suggested, for example, that the R-LOSA suffered from a number of methodological problems in relation to the extensive checklist, which replaced the traditional narrative approach to LOSA observations.

The authors reported that the FAA's original set of R-LOSA observation forms consisted of 41 pages, to be filled in by the observer. De Boer et al. (2013) reduced that to an 8-page checklist and a 12-page observation form. Due to those changes, the researchers were not able to use FAA's standard software, as it could not be modified to suit the new codes or forms. KLM commissioned its own Microsoft Access application to meet its customised requirements.

With the new tools, the authors suggested the method was successful in identifying threats and errors on the ramp. Data showed that, on average, observers witnessed 0.9 external threats and 15.8 errors (compared to 3 or 4 errors per flight for cockpit operations), which suggested that there are higher numbers of errors observed in ramp operations compared to cockpit operations.

The study was able to identify high-frequency threats, such as 'airside speeding', which occurred in 80 per cent of observations, and 'failure to check for foreign object debris', which was cited in 64 per cent of observations (p. 7). The study also found that, on average, 30 per cent of regulations are not complied with during turnaround.

Observers also captured errors leading to injuries, such as ‘failure to use barriers on conveyor belts’, which led to baggage falling on employees and four lost-time accidents (p. 7). Another common error was ‘not checking the hold after unloading’, affecting 69 per cent of observations, ‘setting conveyor height during driving’, which was reported in 62 per cent of observations, and ‘employees walking on running conveyor belts’, which was noted 48 per cent of the time (p. 8). None of the issues identified in the observation process were reported through the airline’s other safety-reporting mechanisms or incident databases, suggesting that the R-LOSA program did produce data that was unique and not available through other sources (De Boer et al. 2013).

In a later paper, De Boer and De Jong (2013) pointed to several issues they experienced with the R-LOSA approach, such as problems with lengthy/elaborate forms and the inflexibility of standardised software. Perhaps more seriously, however, they were concerned that, although the program was effective at identifying threats and errors, the results had not led to development of effective safety interventions. De Boer and Jong noted that, aside from the initial shock at the high number of violations reported, ‘contradictory to what was expected, a very limited improvement in safety performance has been realized over the last twelve months, despite the magnitude of the infringements’ (p. 1).

It is unclear why the data from the R-LOSA program did not generate improvements. This could be related to the quality of the data provided by the program, its interpretation and translation into interventions, the effectiveness of interventions or even the organisation’s openness to change. The reasons for the stagnation in safety improvement described by De Boer and De Jong were not explored further. However, this raises an important point. LOSA is merely a data-collection tool; while the quality of the data produced is important, there are many factors that will determine whether safety improvement will result from a LOSA program.

In 2013, De Boer et al. suggested that the R-LOSA approach has methodological weaknesses. In particular, they criticised the 'questionable TEM framework. (p. 30). De Boer (2013) argues that the TEM model in the ICAO guidance material 'unjustly emphasises threats' (p. 7). Criticisms include the description of threats as necessary precursors to errors and the suggestion that the identification of threats is susceptible to hindsight bias. The authors also suggested that the relationships reflected in the model may not reflect the reality of how threats and errors are experienced and managed in the ramp environment (p. 7). They suggest the TEM model may warrant further investigation and suggest consideration of more simplified models

Based on their own data, De Boer et al. (2013) suggested that errors may initiate with the ramp employees themselves and that the link between threats and errors may be more 'sporadic' than the original ICAO TEM model had indicated (p. 4). They concluded that, if this was the case, 'the identification and resolution of threats may only have a limited effect on safety' (p. 8).

More recently, De Boer (2017) suggested that the TEM theory has a number of flaws. De Boer raises concerns regarding the definition of a threat, suggesting that it lacks a clear, unequivocal description in LOSA. De Boer also questions whether a threat is a really a condition that 'always needs management' and notes that threats are identified 'after the fact' (i.e. with hindsight). He also raises the point that LOSA and TEM assume a normative base for behaviour, and questions whether all variations from a norm should be considered undesirable. De Boer suggests instead that it may be more valuable to assess the gap between 'Work-as-Imagined and Work-as-Done' (De Boer 2017).

This application of R-LOSA to the ground environment in Schiphol airport highlights a number of important considerations for the current research. Previous studies have suggested that context is important when coding behaviour. The use of checklists rather than narratives saves time but raises concerns that important contextual information may be lost. Without the narrative, it is difficult to see how the data verification process is achievable, that is, if the wash team do not have the context to understand the codes assigned. This raises further concerns about reliability of the coding process, if consensus verification is not achievable with checklists. In addition, the difficulties encountered with the TEM model suggest that this concept of human performance requires further review and verification. The study at KLM would seem to indicate that the tools and model may require further development. Furthermore, it will be important to consider how data from programs like R- LOSA can translate into organisational change.

Whilst the method's proliferation outside the cockpit demonstrates confidence among practitioners that it can be translated to other environments, it does not address concerns that the method itself lacks validity. Methodological and theoretical issues surrounding the LOSA method and its underlying TEM model will therefore be reviewed more closely below.

For the sake of clarity when referring to the LOSA methodology, this thesis will refer to the original TEM model described in the ICAO 2002 Doc 9803 and the definitions contained there and further refined in the FAA 2006 advisory circular. It is recognised that since LOSA began in 1999 there have been several interpretations and adaptations to the original TEM model; these are further outlined in Appendix 2. However, for the purposes of discussion and comparison, the LOSA ICAO 2002 version will be used as it is the most commonly cited source document in LOSA implementations.

5.4 Criticisms of LOSA and the Need for New Approaches

Despite its widespread application, LOSA and the TEM concept has also been the subject of some criticism and debate. Some theorists, for example, have questioned its proliferation in the absence of evidence to support its underlying theoretical frameworks or claimed benefits. Angell and Straub (1999) used the term 'consensus authority' to describe the self-sustaining energy that comes from a theory's popularity (p. 184). Dekker argues this consensus authority, rather than any actual evidence that the method is measuring what it claims to measure, underpins LOSA's success (2003). Academic and safety theorists have raised a number of questions regarding the approach that require closer examination. This section will consider the merits of those criticisms and how they might be addressed in the development of a method for ground safety.

The criticisms levelled at the LOSA and its underlying TEM model can be grouped into three broad areas:

1. Concerns regarding validity of the TEM model
2. Criticisms regarding the observation and classification of error
3. Criticisms that LOSA focuses on human error rather than human adaptability.

Each of these will be explored in greater detail below to determine whether they are warranted and, if so, how they might be addressed.

5.4.1 Concerns regarding validity of the TEM model

Dekker (2003, 2007) criticises the aviation industry for adopting TEM and LOSA without testing or challenging its concepts through critical research. Adaptations of LOSA outside the cockpit have adapted the method without further examination of the TEM model or its theoretical concepts. Dekker (2001) describes TEM as an invalid and unproven 'folk model' for directing observation – one that leads to meaningless classifications of human behaviour (p. 3). Whilst is fair to say that LOSA and TEM have not received much academic scrutiny, Dekker perhaps goes too far in suggesting it is invalid. To date, LOSA and the TEM model upon which it is based have neither been validated nor invalidated. TEM and LOSA have merely been proposed as a theory and an approach for collecting data.

When first proposed by Helmreich and his research team, the TEM model was developed to help organise and describe the data collected from observations of pilots' performance. Helmreich et al. (2001) claimed only that the model seemed to fit with the data from cockpit observations and that it appeared to add value when organising the data from their research.

The value referred to could be described as face validity, in that the model appears to make intuitive sense. Some have seen this as an advantage of LOSA. Bove (2002), for example, applauded LOSA's intuitiveness, suggesting that the concept does not require a theoretical background, which makes it easy to understand and, therefore, readily accepted by industry. McLeod (2012), however, suggested that it is 'easy to understand because there is, in fact, no theoretical basis for the model', which he viewed as a serious weakness (AAVPA conference address 2012).

MacDonald (2008) raised concerns about LOSA's 'loose theoretical concepts' (p. 1). Dekker (2003) also criticised the lack of demonstrated validity of the model, suggesting that 'classifications systems like LOSA and the un-validated models upon which they are based, are a form of "self-fulfilling prophecy"', meaning the method seeks out what is described within the model rather than objective phenomena. Dekker (2003) argues:

> The error becomes true… only because a community of specialists has developed tools that would seem to make it appear, and have agreed on the language that makes it visible. There is nothing inherently 'true' about the error at all. Its meaning is merely enforced and handed down through systems of observer training, labelling and communication of the results [aided by] industry's acceptance and promotion (pp. 179–80).

Some criticism of the LOSA approach may stem from a flawed observation study involving air traffic controllers, conducted by Dedale and discussed earlier in this chapter. Hollnagel and Amalberti (2001) suggest that the conclusion of this study 'could very well be that attempts systematically to identify "human error" through observation in practice are doomed to failure' (p. 2). McDonald (2008) also notes:

> When the diagnosis or prediction supported by a theory is wrong, then, while it is not necessarily the case that the entire theory is wrong, it is the case that important fundamentals of the theory must be wrong. We have a case study of the application of TEM – LOSA (threat and error management – line operations safety audit) where the diagnosis supported by this methodology was almost diametrically contrary to the real situation. The reason for this derives from fundamental definitions of error (p. 2).

Rather than being a falsification of an observation approach, such case studies might provide an opportunity to refine what works (and does not work) in human performance observation – such as the importance using of peer observers, retaining the context of the operator, and the inappropriate use of observations for cognitive tasks. Nevertheless, questions regarding the validity of the original TEM model and its classification of error require further consideration.

The application of R-LOSA in KLM by De Boer et al. (2013), for example, suggests that the original TEM model, as published by the ICAO in 2002, over-emphasises threats as necessary precursors to errors. Later studies by LOSA developers Klinect et al. (1999) found that that less than 10 per cent of cockpit crew errors had a threat identified as precedent. This would suggest that the relationships in the model may need revising. The need for refinements, however, need not suggest the whole approach should be thrown out. Rather, LOSA and TEM should be seen as a model and approach offered as a proposal for others to test, refute or refine through further scientific inquiry.

In order to test a model, a set of hypotheses is needed about what could be expected in real-world human performance if the model accurately represents reality. These assumptions can then be validated or refuted through research. Questions for future research might include: Does the presence of threats lead to an increase in errors? Can we determine which errors, in particular, are more likely to lead to other errors or undesirable states? Do all threats and errors need to be managed or are some more important to safety outcomes than others?

Dekker probably goes too far in suggesting that we should abandon the method altogether. Rather, future research should seek to test and refine the model and its assumptions. When adapting LOSA for use in ground safety, new approaches should seek to test and strengthen the construct validity of the TEM model and LOSA method where possible.

5.4.2 Criticisms of the observation and classification of error

This criticism is related to what is considered an error, how it is labelled and whether or not it is observable. LOSA error classifications have been criticised for confusing cause and consequence, resulting in labels that are not mutually exclusive (Dekker 2001; Dekker 2003). Dekker (2003) notes: ‘LOSA contains categories of manifestations of error (such as communication errors) as well as causes of error (such as proficiency problems), thus it mixes causes and consequences’ (p. 97). Isaac et al. (2002) described how taxonomies that are ‘simply a collection of items, which are not necessarily comprehensive and whose categories are not mutually exclusive, lead to ambiguous or fuzzy results or, worse, [an] erroneous and misrepresentative aggregation of errors’ (p. 2).

Using LOSA codes, a pilot’s decision to navigate through adverse weather could be labelled in three different ways: as an operational decision error (a misjudgement of the conditions); a procedural error (unfamiliarity with the rule); or even the result of intentional non-compliance (deliberately violating the rule). Rasmussen’s (1982) work on human error taxonomies argued that classification of human error should be based on a single organising principle, such as describing either what went wrong, how it went wrong, the effect upon the external task or why it happened; and that the use of more than one organising principle leads to confusion (p. 319).

The most appropriate organising principle to choose may depend on the data collection method. Rassmussen argued that the mechanisms of malfunction, or causality, are often not directly observable, which would make LOSA observations unsuitable for this task. Shorrock and Kirwan (2002) agreed that the causes of errors involve nonvisible cognitive processes that cannot be deduced from observation alone. However, LOSA terms such as decision-making error and intentional violations imply that the observer is able to deduce such cognitive mechanisms through the observation alone, which is difficult to justify. MacDonald (2008) asked, for example, ‘Is intentionally not following a procedure an example of a violation, productive sense-making or both?’ (p. 2). Unless a subject voices his or her motivation for breaking a rule or deviating from standard procedure, MacDonald added, ‘all that can be known, for certain, is that the rule was broken’ (p. 2).

Furthermore, the LOSA coding process also requires observers to link threats, errors and consequences together in a way that implies a causal relationship in the TEM model. When a threat or error is observed, the observer must determine whether it is ‘consequential’; that is to say, whether it is linked to another error or to an undesirable aircraft state. However not all cause and effect relationships are observable, especially if they are separated by time and space. Speculation about the causal relationships between these factors is what Hollnagel and Amalberti (2001) describe as the problem of ‘backwards causality’, where observers reason backwards from the effect to determine the cause (p. 3). Several theorists have argued that this process leads to conclusions that are not logically valid (Wason and Johnson-Laird 1972; Tversky and Kahneman 1984; Hollnagel and Amalberti 2001).

The benefit of the observational approach is its supposed objectivity. Therefore, to preserve objectivity, coding should be restricted to observable phenomena. To avoid confusion, categories in a LOSA method should be based on one determining factor, such as the observable external manifestation of the error. Any future adaptation of the LOSA methodology will therefore require redefining of error into more reliable and mutually exclusive coding categories that are restricted to describing observable behaviour. Associations between threats and errors may be better investigated through post hoc statistical analysis, rather than observer interpretation.

5.4.3 LOSA focuses human error rather than human variability in context

Several theorists have argued that approaches like LOSA are overly focused on humans as a source of failure, rather than on their resilient capacities (Senders and Moray 1991; Woods et al. 1994; Hollnagel 1983). Others have argued that the even the use of the term 'error' should be avoided altogether (MacDonald 2006). Hollnagel and Amalberti (2001), referring again to the Dedale air traffic control study, suggest that what is labelled as error is not just dependent on who makes the judgement but on whether the outcome is good or bad.

MacDonald (2006) argued that a variation from written procedures (technically an error in LOSA) is not necessarily problematic and can sometimes be advantageous. When viewed in their context, errors could, in fact, be very sensible adaptations (Dekker and Hollnagel 2001). Others have argued that error classifications take errors out of their local context, which means the reasons for the error from the human perspective are lost (Angell and Straub 1999; Vaughan 1996; Snook 2000; Dekker 2001). Dekker claims, for example that 'Error classification disembodies data. It removes the context that helped produce the error in its particular manifestation' (p. 98).

Rasmussen (1982) was one of the first to argue for the importance of focusing instead on human adaptability. He suggested the variability of humans is what allows us to adapt to peculiarities in system performance, which he believed was ‘the very reason for having people in a system’ (p. 313). Reason (2008), in his more recent works, shifts focus away from ‘unsafe acts’ to consider the ‘heroic recoveries’ that humans have made possible. Dismukes (2009) also recognised that for every accident attributed to human error, vast numbers of accidents are averted by the skilled intervention of human operators. These averted accidents receive little attention, though we could learn much from studying them, if the data were available (p. 680).

For Dekker and Hollnagel (2003), ‘[t]he point in learning about human error is not to find out where people went wrong, but is to find out why their assessments and actions made sense to them at the time, given their knowledge, goals, tools and limited resources’ (p. 98). They suggest error classification takes errors out of context and ‘disables understanding’ (p. 98).

Proponents of LOSA would disagree with this assessment. The original LOSA concept was designed to observe everyday behaviour in its natural environment – within its natural context. Indeed, proponents of LOSA have argued that the recording of context in normal operations is a particular strength of the methodology (Maurino, 2005). With its focus on normal operations, LOSA records how humans typically cope with everyday threats and irregularities to produce mostly safe outcomes. When following ICAO guidance LOSA uses experienced peer observers so they can contextualise the behaviour they observe. Observers record as much contextual detail as possible during an observation to help make sense of decisions that were made at the time during the coding phase. Context is also vital in the data verification stage where observers use the narratives to agree via consensus on the codes discussed.

Context is also important in LOSA when interpreting how threats and errors are managed. In his original thesis, Klinect (2005) downplays the importance of quantifying the prevalence of threats, errors and UOS and stresses that it is the management of these phenomena that should be of particular interest (p. 78). Klinect explains how the narratives provide an ‘operational context drill down’, to ‘fill gaps in understanding’ and provide for deeper layers of analysis (Klinect p. 85) than is available from the data alone.

So it is important to understand, firstly, when behaviours vary from what is expected, the context in which this behaviour made sense at the time, and also the management of those conditions and behaviours in normal operations. It is also important to understand the safety outcome of the behaviour. Hollnagel and Amalberti (2001) argue that whether an action is seen as being correct or incorrect depends whether the result is desirable ‘because the distinction applies to the outcome rather than the action itself’ (p. 3).

LOSA observations are able to record all the occasions where crew vary from the procedures, the context in which these actions occurred and how these conditions and behaviours are managed and the outcome. Determining if the action taken was successful depends on the outcome. In view of the concerns about LOSA terminology it is perhaps time to reconsider the labels used and whether they infer humans are a cause of failure or a source of success.

McDonald (2008) suggests that models should avoid using basic theoretical terms that are not value neutral (e.g. error, violation) (p.2). The term error be replaced with a neutral description that simply describes an observable action such as a deviation from an expected behaviour or a variation from a procedure, for example.

Whilst there may indeed be a problem with the way the behaviour is classified. All procedural variations are of interest, irrespective of the label given to them or the reasons for their occurrence, because of their potential to impact safety outcomes. This is because organisational systems mostly rely on humans in the system to perform tasks in predicable ways, which are defined in a procedure. Flexibility of response can be useful but so are procedures, in order to achieve predictable outcomes. This is not to say that procedures are always right; rather, it is important both to have defined ways of performing tasks and to know when procedures are not working as intended, so they can be improved.

When developing new observation methods for ground safety it may be advisable to redefine error as a 'variation from standard', to more accurately describes the phenomenon observed. The standard in this case, should be defined by the organisation and could be defined in the current operating manual or procedure. Observations could note when procedures are varied, how this was managed and the outcome that followed. Statistical assessment of the behaviours most associated with successful and unsuccessful outcomes would provide valuable feedback for the safety of the system.

When developing new observation methods for ground safety it will be important to note when a task was not carried out as described in the procedure, how this was managed and the outcome that followed. It will also be important to retain the rich, contextual data that is captured in narratives so that these actions and outcomes can be understood in context. Statistical assessment of the behaviours most associated with successful and unsuccessful outcomes, would provide valuable feedback for the safety of the system.

5.5 Summary and Implications for the Current Research

5.5.1 Lessons learned from previous applications

Several lessons can be learned from the application of LOSA in flight operations to date. Whilst there is no evidence that flight crew CRM would be beneficial for ground crews, the use of an observation method could provide new evidence to demonstrate which behaviours produce safe outcomes for future training targets. Ideally a new method should statistically identify successful and unsuccessful behaviours from observation data.

Furthermore, data from observing normal operations provides feedback about how people interact with equipment and procedures that may be useful for improving the usability of procedures and equipment on the ground. The use of a model of human performance, like TEM, may be helpful for assigning common terms, which is useful for both proactive data collection and incident investigations, to quantifying how often some behaviors occur in normal operations and for sharing and comparing incident data across the ground industry sector.

Whilst the link between learning from LOSA and safety performance cannot yet be objectively established, it is clearly seen as valuable to the aviation industry. LOSA has considerable face validity and perceived value among both industry insiders and regulators, which gives it an advantage by encouraging acceptance of a new LOSA if these perceived benefits could be transferable to the ground handling sector.

LOSA has already been adapted for several other industrial settings, which suggests the method is transferable. From these examples, however, it is possible to identify a number of important considerations. LOSA approaches seem to works best in highly procedural environments and are less suited to observing cognitive or unobservable tasks as identified by early studies into air traffic control. (Henry 2003 Hollnagel and Amalberti 2001). Another conclusion from the discussion above may be that LOSA adaptations should develop their own domain-specific, standard taxonomies that make sense in the operational context – as did those developed for CORS in rail, ANTS in healthcare and R-LOSA on the ramp. These may be derived from a number of sources such as observation, task analysis and input from subject matter experts.

Lessons from CORS, put forward by MacDonald et al. (2006), suggested that the principles of the methodology were transferable outside of the aviation industry but that standard CRM behaviours developed with pilots in mind may not be. Rather they suggested using the data from observations to identify successful and unsuccessful behaviours in each domain. Those insights can then be applied to future training and used to evaluate behaviour change through repeat observations. The importance of maintaining domain-specific context was also highlighted when selecting observers, with Hollnagel and Amalberti (2001) suggesting that external observers lack the context through which to interpret behaviours effectively. Peer observers with knowledge and experience of the operational context are, therefore, an important asset when interpreting target behaviours.

The healthcare-related studies also highlight the importance of a standardised taxonomy that serves as a tool for agreeing on common terms, data-collection and database entry (Sanders and McCormick, 1993) and the importance of providing extensive training to achieve inter-rater reliability among observers (Fletcher et al 2003). Methods that enhance inter-rater reliability (e.g., training, calibrating observer coding through video, data verification) will, therefore, be important aspects in the development of a new method.

The R-LOSA tools developed concurrently to this research project indicate the need and an appetite for such programs in the ground safety sector. The difference between the original LOSA tools and the R- and M-LOSA tools, however, also highlight some important issues. There may be numerous advantages to maintaining the narrative form of observation recoding from LOSA, rather than the checklist approach. Narrative provides important context to interpret behaviours during coding and more importantly during the coding verification process to ensure the consistent application of codes which is vital for reliability.

5.5.2 Implications for the development of a tool for ground handling

When consideration of the LOSA method for ground safety it would appear to offer a number of advantages such as provision of a model for assigning common terms for proactive data collection and incident investigations; the ability to determine which behaviours on the ramp produce safe or unsafe outcomes for future training targets; useful feedback about how ramp crew interact with equipment and procedures.

LOSA methods seem to work best in environments with observable tasks, which may make it suitable for the ramp. The method already enjoys credibility and face validity in aviation and it would appear that the method is transferable. Adapting a LOSA method for the ground would require codes specific to the ground handling domain with peer observers familiar with ramp operations. Even with experienced observers, extensive training in the use of codes will be required. It would seem advisable to retain the narrative recording format over the checklist for context improved interpretation, verification and reliability. There are still issues concerning the validity of the model and methods that will need to be overcome. A more robust and structured evaluation process is needed.

Most of the adaptions of LOSA appear to have incorporated the TEM model and theory of LOSA without change or consideration for its validity. The exception to this was the KLM implementation of R-LOSA tools, which was found to have weaknesses in methodology and raised concerns about the 'questionable TEM framework' upon which it was based (De Boer et al 2013). LOSA has face validity and perceived value that make it worth pursuing.

The LOSA method and TEM model have remained largely unchanged since the early 90's with only minor modification (See Appendix 2) despite the development of new concepts in safety science. It is perhaps time for a review of those concepts in view of contemporary theories. Hollnagel and Amalberti (2001) have too hastily concluded that methods that identify human error through in-practice observation are 'doomed to failure' (p. 2). Conclusions that the entire approach should be rejected are unwarranted. Rather what is needed is a more structured analysis to determine LOSA/TEM's strengths and where improvements are needed. With this understanding it may be possible to develop a new method that builds on LOSA's strengths but avoids some of its methodological weaknesses.

A more robust and structured evaluation process is needed. Structured methods for evaluating the effectiveness of tools are available and have been applied in a number of different domains. Evaluation frameworks for example, can assess a tool against a number of defined criteria, to determine its methodological strengths and weaknesses. The next chapter proposes an evaluation framework that may be suitable for the evaluation of LOSA's effectiveness as a data collection tool and considers the most suitable criteria for this purpose.

Chapter 6: Developing an Evaluation Framework

6.1 The Need for an Evaluation Framework

Whitefield, Wilson and Dowel (1991) argue that 'evaluation is an integral part of any development process' (p. 65). When developing a new tool it is essential to evaluate the resulting product to determine its value or its success in meeting the tool's original objectives. One way to assess the success of a tool is through the application of an evaluation framework. An evaluation framework is usually a set of criteria against which a product, such as a tool or methodology, can be assessed. Mitchell et al (2009) outline how evaluation frameworks can provide clearly defined characteristics or attributes in a common instrument, upon which different tools can be compared and contrasted. The application of the framework illuminates the strengths and limitations of data-collection instruments, (Mitchell et al., and 2009). Similarly, Bott, Guedes and Claramunt (2010) explore how evaluation frameworks can help developers of new tools to design, and subsequently evaluate, their methods by helping to:

- define the tool's development goals and objectives,
- articulate the elements that affect the tools success,
- define the questions that need to be answered, to gauge success of the tool, and
- Define the information needed to determine if the expected objectives and outcomes for the tool have been accomplished.

There are several examples of evaluation frameworks that have attempted to compare and contrast tools against a range of defined criteria (e.g. Mitchell et al., 2009, The Centre for Disease Control and Prevention, 2001, 2004, Kirkpatrick 1977, 1979).

In the area of human performance, one of the most comprehensive evaluation frameworks was developed by Kirwan (1992, 1996, 1998) to qualitatively assess the relative merits of human error identification (HEI) tools from the human factors practitioners' perspective. Kirwan was interested in comparing HEI tools for use in human reliability assessment (HRA) and probabilistic risk assessment (PRA), (discussed in Chapter 4). These share some common objectives and features with LOSA, such as the identification and classification of human actions based on a model of human performance. The model is used to structure and code the data collection and the interpretation of results.

Barber and Stanton (1996) provide an example of how Kirwan's original principles have been adapted and applied for the purposes of evaluating product development. In this example, Barber and Stanton wanted to compare different HEI tools to predict errors in the use of a ticket vending machine. Barber and Stanton referred to Kirwan's original criteria, but expanded and adapted these concepts where necessary for their purposes. In the current research, both Kirwan's (1992b) and Barber and Stanton's (1996) criteria will be considered to develop a bespoke evaluation framework suitable for the purposes of the current research. The adapted criteria will then be discussed in relation to LOSA as it is today, to identify features of LOSA that should be retained and those that should be modified prior to its application in ground safety.

6.1.1 Background to Kirwan's evaluation framework.

Kirwan's early work (Kirwan 1992a) set out to review the available techniques for human error identification (HEI), highlighting these techniques' strengths and weaknesses. He came to the conclusion that a good tool should meet three important criteria: 'comprehensiveness' in identifying possible errors, 'usefulness' for predicting and reducing error, and 'documentability' to ensure the tool provides a useful record of data over time. In a second paper (Kirwan 1992b, Part 2), Kirwan went on to evaluate several HEI tools against an expanded range of criteria that also considered the tools' practical usefulness (Kirwan 1996, p. 311). He described the list of criteria as 'an exhaustive and formal set of criteria to determine which techniques are particularly useful (Kirwan 1998, p. 168). In total, seven criteria were applied (1992b, p.274). These criteria were:

- **Comprehensiveness**: accuracy and completeness for identifying errors,
- **Consistency**: the degree to which different assessors produce similar results,
- **Usefulness**: in terms of providing diagnosis and identifying error reduction solutions,
- **Resources**: in terms of the time, effort and expertise needed to apply the technique,
- **Acceptability**: the availability of the tool for general use and its application to date,
- **Auditability**: this criteria includes the previous criteria of documentability and refers to the documenting of results for audit purposes, and
- **Theoretical validity**: whether the tool is based on a valid model of human performance.

The evaluation compared and rated the techniques as 'high', 'moderate' or 'low' for each of the evaluation criteria defined. It is interesting to note that most of the human error techniques evaluated were rated as moderate or low against the criteria defined by Kirwan, suggesting that new approaches might be welcome. Nevertheless, the criteria themselves have become a useful mechanism for evaluating human performance evaluation tools in general.

These seven basic principles, for example, were also applied by Barber and Stanton (1996) with some subtle differences. They applied their criteria to assess the use of HEI techniques as a possible alternative to observation studies in order to predict errors in the use of a ticket vending machine. Barber and Stanton found that HEI techniques and observation techniques were comparable in terms of being able to identify similar errors in consumer products.

These seven principles applied by Kirwan and Barber and Stanton may also prove useful for evaluating which features of LOSA should be included in a ground safety human performance tool and which should be adapted.

6.2 Criteria for the Proposed Evaluation Framework

Both Kirwan's and Barber and Stanton's definition of each criterion will be compared and contrasted in more detail below and used to develop a set of criteria to be applied in the current research.

6.2.1 Comprehensiveness

For Kirwan (1992b), a primary criterion of any error identification tool is 'comprehensiveness', particularly with respect to identifying critical or serious errors. Kirwan described comprehensiveness as the degree to which the technique is able to identify significant errors and the method's breadth in dealing with all forms of error (1992, p. 372). Kirwan also thought a comprehensive model should also be able to identify influencing and performance shaping factors in the environment. Kirwan rated comprehensiveness favourably if it was able to identify different types of skill rule and knowledge based behaviours.

Similarly, Barber and Stanton (1996) described 'comprehensiveness' as the 'accuracy with which errors are uncovered or concurrent validity' (p. 120), which they related to 'the number of error types predicted to those that were observed during the actual use of an artefact', rated on a scale of 1 (poor) to 5 (good).

For the purposes of the current research, Barber and Stanton's definition would be too narrow; therefore, for the purposes of the current research a criterion more closely related to Kirwan's criteria should apply. Comprehensiveness will consider the completeness with which an observation tool can describe the full range of human behaviour observed, as well as observable factors in the environment that influence those behaviours. For simplicity, this can be rated on a scale of 1 (poor), 2 (moderate) or 3 (good).

6.2.2 Consistency

Kirwan described 'consistency' as 'a function of how well structured the technique is, so that it is not open to many different interpretations' (p. 372). Kirwan argued that less-structured, open-ended techniques would lead to results that were 'highly assessor-dependent' (p. 375), whereas more structured approaches lead to different assessors applying the same categories and assumptions. Kirwan rates consistency as 'low', 'moderate' or 'high': low indicates a relatively open-ended technique, moderate suggests that the assessor has flexibility within a detailed framework, and high indicates that the tool is highly structured and likely to lead different assessors to identify the same things, given the same information and assumptions.

Similarly, Barber and Stanton (1996) described consistency as 'the reliability with which different assessors can use the technique for different domains' (p. 120), which they rated on a scale of 1–5 with 1 being poor consistency and 5 good consistency. Kirwan, and Barber and Stanton, were both referring to the tool's ability to produce consistent and reliable data and inter-rater-reliability.

For the proposed evaluation framework, the criterion of consistency will therefore refer to the degree to which the technique is structured to achieve reliability so that different assessors describe and categorise the same observable phenomenon in the same way. Consistency for this criterion will be rated on a scale of 1 (poor), 2 (moderate) or 3 (good).

6.2.3 Usefulness

Kirwan (1992b) described 'usefulness' as 'the degree to which the technique generates error-reduction mechanisms irrespective of whether these are based on analysis of root cause or not', and is therefore useful for achieving the overall aim of such methods. Kirwan also considered whether the method provided insight and 'diagnoses into the causes of the error in order to generate improvement initiatives' (p. 168). For example, 'usefulness' is considered to be low if the method is not concerned with error reduction, moderate if the technique is capable of error-reduction, or high if generating effective error reduction is a primary focus of the approach.

In contrast, for their application Barber and Stanton (1996) were more interested in ‘the perceived usefulness’ of the technique for predicting errors, and its ‘perceived usefulness to stakeholders’ or ‘face validity’ (1996, p. 12) from the perspective of all those who use the tool or its outputs, which they rated on a scale of 1–5. Barber and Stanton believed that face validity was the stakeholder’s subjective response to the method’s ‘comprehensiveness’ and its theoretical validity.

Since the purpose of collecting data about human factors and safety in ground operations is to drive safety improvements, Kirwan’s criterion of usefulness will be important for any adaptation in ground operations. However, Barber and Stanton’s understanding of perceived usefulness to stakeholders is also important. It is therefore proposed that, in the current research, ‘usefulness’ should refer to the priority given to generating useful improvement strategies and the perceived usefulness of the tool and its outcomes to industry – rated on a scale of 1 (poor), 2 (moderate) or 3 (good).

6.2.4 Use of resources

Kirwan refers to this as the ‘burden of resources’ needed to conduct the technique. Kirwan specifically differentiates three types of resources: (1) those required to apply the technique in terms of assessor/expert time, rated as high medium or low; (2) whether training is required to use the system, rated as yes or no; and (3) whether the method requires an expert panel of task domain experts, rated as yes or no.

Similarly, Barber and Stanton (1996) are also interested in the resources needed, such as time and expertise. They rate ‘use of resources’ more simply; on a scale between 1 (poor), meaning required time and expertise is high, and 5 (good), indicating low use of time and expertise. The proposed criterion for the current research should also consider the cost of implementing the tool in terms of the resources required – i.e. time, training and expertise. For simplicity, this can be rated as 1 (poor), suggesting that burden to industry is high, 2 (moderate), suggesting a moderate resource burden, and 3 (good), to indicate low use of time and resources.

6.2.5 Acceptability

Kirwan defined acceptability in two ways; ‘usage to date’ and ‘the availability of the technique’ (p. 377). Kirwan also believed that the techniques that were most acceptable to users would be those with which they were most comfortable based on previous use, and those that were readily available to all (p. 373). Kirwan rated acceptability as high, medium or low, with a high rating indicating the method was widely available and had received extensive usage.

In contrast, Barber and Stanton (1996) rated acceptability in relation to the results the method produced, defined as ‘the degree to which all parties concerned can accept the outcome – which, they noted, ‘seems analogous to the concept of content validity’ (p. 120), presumably because stakeholders’ confidence in the method is dependent upon the underlying model upon which it is based. Barber and Stanton rate acceptability on a scale from 1 to 5, where 1 indicates poor acceptability for the tool’s outcomes and 5 indicates high acceptability.

Since any new tool developed will be a prototype, availability and previous usage will not be relevant. Therefore, it is proposed that ‘acceptability’ here should refer to the degree to which all stakeholders accept the underlying model of human performance and thus the results generated by the method. Acceptability of the method can be rated on a scale from 1 to 3, where 1 indicates low levels of acceptability of the results to stakeholders and 3 indicates high acceptability.

6.2.6 Documentability/auditability

Kirwan (1992b) refers to documentability as ‘the degree to which the technique lends itself to auditable documentation’ (p. 377). Kirwan judges this to be low if the technique is difficult to document, moderate if it provides sufficient documentation to be repeatable, and high if all assumptions are recorded and documentation is available for future system operations and to facilitate future periodic assessments.

Barber and Stanton (1996) refer to a similar set of features but define this criterion instead as auditability, 'the degree of transparency and traceability within the process with which the error identification can be transferred from the analyst to the client' (p. 120), which they rated on a scale between 1 (suggesting poor auditability) and 5 (suggesting good auditability).

Kirwan and Barber and Stanton were both concerned with the ability of a method to record and document results. Barber and Stanton's application, however, is too narrow in considering only how information is recorded and transferred to the developer of new products.

Kirwan's criterion is therefore more relevant to evaluate a tool developed for the documenting and periodic assessment human performance. For the proposed evaluation framework, Kirwan's criterion of documentability will be applied to assess how well the results can be documented and repeated for the purposes of periodic audits to provide a baseline of performance and drive, and to allow continuous improvement. This will be rated on a scale of 1–3, where 1 indicates poor documentability and 3 indicates that the results are well documented for the purpose of periodic assessments.

6.2.7 Theoretical validity

Kirwan considers theoretical validity in several different ways. Firstly, he assesses 'whether or not the technique is based on a model of human performance' (p. 375). Theoretical validity is considered to be low if the tool is simple classification-based taxonomy, moderate if the technique makes reference to a model of human performance, and high if it is the direct interpretation or embodiment of a model of human performance (1992b, p. 376). Kirwan also considers whether the model is limited to describing what happened using external error modes or (EEM) and psychological error mechanisms (PMs) to describe how the operator failed. He also noted whether the tool recorded performance-shaping factors (PSFs) or external factors that influenced the operator's performance. Although not a rating per se, Kirwan denotes each method as either able to describe 'ALL' of the above or the acronyms (EEM/PS or PSF) to describe how the tool identifies error.

Barber and Stanton (1996) are not concerned with these different levels of error description but they are concerned with the use of an underlying model. They consider theoretical validity not just in terms of whether a model is used, but whether that model is valid. Specifically, they are concerned with the 'underlying model of human activity that drives the approach' and 'the ability of that model to coherently describe the activity observed' (p. 120), which they refer to as 'construct validity'. They rate theoretical validity on a scale of 1–5, where 1 indicates poor theoretical validity of the underlying model and 5 suggests a method with high validity in terms of its underlying model.

Observations methods can only observe external error modes (EEM) and observable performance shaping factors in the environment (PSFs). The ability of the method to fully describe these features is covered by the first criterion, comprehensiveness. Since it is not possible to document psychological mechanisms (PMs) using an observation method, this distinction is not useful for the current research. More important for the evaluation of theoretical validity, however, are Barber and Stanton's considerations of whether the underlying construct or model is valid. In addition, this criterion should go one step further to consider the validity of the data collection methodology employed. Theoretical validity overall will be rated on a scale of 1–3, where 1 indicates a method and model with poor theoretical validity, 2 indicates moderate theoretical validity and 3 indicates a tool with convincing theoretical validity in terms of the underlying model and the method used to collect data.

6.3 Summary of the Proposed Evaluation Framework

Table 1 (next page) summarises Kirwan's original evaluation criteria (column 1) as well as Barber and Stanton's adaptation of Kirwan's criteria I (column 2). The final column summarises the final adaptation of these two definitions adapted for application in the current research. The criteria in column 3 will first be applied to the LOSA methodology to determine which features of the method should be retained and which should be adapted. LOSA will be judged on a scale of 1 (poor), 2 (moderate) and 3 (good) for each criterion.

6.4 Applying Evaluation Criteria to LOSA to Identify Development Needs for a New Tool

Having outlined in Table 1 the seven proposed criteria to be applied for the development of a new tool, the next section will assess LOSA using these criteria in order to determine which features of the LOSA should be maintained and which should be modified. For each criterion, a rating for LOSA is given and suggestions for modification are made where necessary. The aim is to identify a set of development needs for a new ground safety tool, which will be summarised below.

6.4.1 Comprehensiveness of LOSA

According to the new Evaluation Framework, LOSA would score a 2 for comprehensiveness. LOSA's observation method involves recording a complete 'narrative' of the behaviours observed in the cockpit, as described in Chapter 5. The narrative records threats (similar to performance-shaping factors) as well as the human responses, which can include errors or behaviours that manage threats and errors. In this sense, LOSA could be said to be high in comprehensiveness. However, because of the issues regarding the error taxonomy described previously, LOSA may lack necessary and sufficient descriptions for human behaviour.

Table 1: Summary of Kirwan's criteria, Barber and Stanton's adaptations, and proposal for application to current research.

Kirwan's original Criteria (1992b)	Barber and Stanton adaptations (1996)	Proposed evaluation criteria for development of a new tool
Comprehensiveness – the degree to which the technique is able to identify significant errors and the methods breadth in dealing with all forms of error (rated as low/moderate/high).	**Comprehensiveness** – the accuracy with which errors are uncovered, also known as concurrent validity. In their study comprehensiveness was the number of error types predicted to those which were observed during actual use of an artefact (rated on a scale of 1 (poor) to 5 (very good)).	**Comprehensiveness** – the completeness with which an observation tool can describe the full range of human behaviour observed as well as observable factors in the environment which influence those behaviours (rated on a scale of 1 (poor) 2 (moderate) 3 (good)).
Consistency – a function of how well structured the technique is so that it is not open to many different interpretations or the degree to which different assessors can use the technique in the same way to produce similar results (rated as low/mod/high).	**Consistency** – the reliability with which different assessors can use the technique for different domains (rated on a scale of 1 (poor) to 5 (very good).	**Consistency** – the degree to which the technique is structured to achieve reliability so that different assessors to describe and categorise the same observable phenomenon in the same way (rated on a scale of 1 (poor) 2 (moderate) 3 (good)).
Usefulness – the degree to which the technique can be applied in order to generate error reduction mechanisms (rated low (method is not concerned with error-reduction), moderate (method is capable of error-reduction), or high (error-reduction is a primary focus of the approach)).	**Usefulness** – the perceived usefulness of the technique in predicting errors or face validity. Barber and Stanton suggest this is a subjective response to the criteria of comprehensiveness and theoretical validity (rated on a scale of 1 (poor) to 5 (very good)).	**Usefulness** – priority given to generating useful improvement strategies and the perceived usefulness of the tool and its outcomes to stakeholders (rated on a scale of 1 (poor) 2 (moderate) 3 (good)).
Resource Usage – the demands of the technique in terms of the resources needed to implement the tool in terms of expert time (low/mod/high), whether an expert panel is needed (y/n) and whether training is required (y/n).	**Resource usage** – factors as time, expertise and resources in order to use the technique (rated on a scale of 1 (poor) to 5 (very good)).	**Use of Resources** – the cost of implementing the tool in terms of training expertise, time and resources (rated 1 (poor, suggesting that burden to industry is high), 2 (suggesting a moderate resource burden) or 3 (good, to indicate low use of time and resources)).
Acceptability – Kirwan defined acceptability in two ways 'Usage to date' and 'availability of the technique' (p. 377). Rated high, medium or low (with high indicating the method was widely available and had received extensive usage) with a yes/no rating to indicate whether the tool was available to industry in a form ready for general use.	**Acceptability** – the degree to which all parties concerned can accept the outcome or results based on the tool, which Barber and Stanton considered analogous to the concept of content validity (rated on a scale of 1 (poor) to 5 (very good)).	**Acceptability** – the degree to which all stakeholders, can accept the underlying model of human performance and therefore the results generated by the method (rated as 1 (poor) 2 (moderate) 3 (good)).

Kirwan's original Criteria (1992b)	Barber and Stanton adaptations (1996)	Proposed evaluation criteria for development of a new tool:
Documentability/ Auditability – the degree to which the technique lends itself to auditable documentation (p. 377) (rated low (the technique is difficult to document), moderate (provides sufficient documentation to be repeatable) or high (all assumptions are recorded and documentation is available for future system operations and to facilitate future periodic assessments)).	**Auditability** – the degree of transparency and traceability with which the error identification process can be transferred from the analyst to client (rated on a scale of 1 (poor) to 5 (very good)).	**Documentability** – how well the method and results are documented to allow periodic audits to provide a baseline of performance and drive continuous improvement (rated 1 (poor) 2 (moderate) or 3 (good)).
Theoretical Validity – whether or not the method is based on a model of human performance (rated low (the tool is simple classification-based taxonomy), moderate (the technique makes reference to a model of human performance), or high (a direct interpretation or embodiment of a model of human performance)). Kirwan also considers whether the model is limited to describing what happened (external error modes), or whether it can also interpret psychological error mechanisms (such as how the operator failed internally)and performance-shaping factors (external factors which influenced the operator's performance)	**Theoretical/Construct Validity** – whether or not the method is driven by a model of human activity and whether that model coherently describes the activity observed (rated on a scale of 1 (poor) to 5 (very good)).	**Theoretical Validity** – the validity of the underlying theory or construct of human performance on which the tool is based as well as the validity of the data collection methodology employed (rated as 1 (poor), 2(moderate) or 3 (good)).

When developing a tool for ground operations, the tool must be able to describe the full range of human behaviours encountered in ground operations. This requires the development of a comprehensive range of domain specific codes to describe the performance shaping factors (or threats) in the environment. The tool will also need to reconsider the way errors are described to be able to fully and comprehensively describe all the possible behaviours observed. This issue is related to the consistency of coding, which is discussed below.

6.4.2 Consistency of LOSA

Using the developed criterion, LOSA would score a 2 as a rating for consistency. The LOSA tool is based on some highly structured processes, such as ten specific 'operating characteristics' in the LOSA Advisor Circulars (ICAO 2002, and FAA 2006). These provide structured guidance to set up a LOSA program to ensure that the data collection process is robust. They include processes for the recruitment and training of observers, for the use of observer tools and databases and for data-verification processes, as described in Chapter 5. These processes are intended to encourage naturalistic behaviour and to maintain the consistency of LOSA programs across audit periods and between different airlines. The apparently successful application of the LOSA method outside of the cockpit also suggests the structure of the method allows the method to transferred and usable between domains.

This could lead to the conclusion that LOSA is high in consistency; however, as discussed previously, the LOSA method has been criticised for its problematic taxonomy, which confuses causes and consequences and lacks mutually exclusive descriptions of phenomena. This may result in poor inter-rater reliability and reduce the reliability of the results. For this reason, the consistency rating is reduced from good to moderate.

Adapting LOSA for ground safety should seek to improve both the reliability of the method and the consistency of its application. Nunnally (1978) argues that inter-rater reliability can be improved by making the coding as unambiguous as possible and the rules for coding as explicit as possible. Restricting the coding to observable behaviour should therefore improve reliability. In addition, a new method should retain the process of data verification or washing process. Although this does not improve inter-rater reliability of coding, it can help to improve the consistency of results by gaining consensus on the codes applied.

6.4.3 Usefulness of LOSA

Using the developed criterion, LOSA would score a 2 for usefulness. Although the identification of error-reduction mechanisms is one of the primary objectives of LOSA, critics of LOSA, such as Dekker and (2001), have suggested that the way in which LOSA classifies and counts errors means it is not useful and even 'illusory' (Dekker 2003). Usefulness, therefore, depends upon which stakeholders' perceptions are given most credence. Overwhelmingly, industry stakeholders have perceived the usefulness of LOSA outcomes as very good. This is particularly the case for airlines that have implemented LOSA programs. LOSA has been credited with nine uses or benefits (FAA 2006 and ICAO 2002) which are outlined below with supporting claims from industry. However, the evidence to support these claims is largely anecdotal and comes from airlines or industry groups rather than from published studies. Nevertheless, the claimed benefits are still an indication of LOSA's perceived usefulness to industry and regulators. Perceived uses or benefits of LOSA include:

- Identifying threats in airlines' operating environments,
- Identifying threats and errors from within the airlines' operations,
- Assessing the effectiveness of training,
- Checking the quality and usability of procedures,
- Identifying design problems in the human-machine interface,
- Understanding pilots' short-cuts and workarounds,
- Assessing safety margins,
- Providing a baseline for organisational change, and
- Providing a rationale for the allocation of resources.

LOSA also enjoys high perceived usefulness of its outcomes or error reduction mechanisms. LOSA has been credited in uses such as continuously improving Crew Resource Management training (Thomas 2004; Lauber 1987; Wiener, Kanki and Helmreich 1993); improving checklists and procedures (Graeber 2006; Boorman 2001; Holder 2004); improving the design of the aircraft HMI (Tesmer 2002; Greaber 2006); reducing unstable aircraft approaches (Gunther 2002, reported in Klinect 2005); and improving safety outcomes (Klinect et al. 2003; Veilette 2008 and Gunthur 2002; Craig, 2006).

Although the claims made above cannot be independently validated, perceived benefits such as these have spurred the popularity of the LOSA in the aviation industry and beyond. The credibility and face validity of the method within the aviation industry is therefore a key strength that should therefore be retained and built upon in the development of a new tool for ground operations.

6.4.4 LOSA's use of resources

According to the evaluation criteria LOSA would score a 1, due to its high burden of resources. When considering time, training costs and use of expertise, LOSA is considered costly in all three areas. Resources are needed to apply and set up the technique in the first instance and there is considerable training needed to use the coding and data-verification tools as described in Chapter 5. Observations involve taking personnel (usually pilots) offline to conduct many hours of observations. Further expertise is also required to analyse and interpret LOSA results (ICAO 2006).

Steckel (2014) notes that LOSA programs are most often found in large air carriers that have considerable 'human and monetary resources' to support them, but notes that, for these carriers, the perceived benefits of implementing LOSA seem to justify the investment (p. 84). The proliferation of LOSA as a methodology would seem to suggest that it is perceived by any to have value that justifies the initial investment. In contrast, no HRA methods have taken hold in aviation (as discussed in Chapter 4). However, no cost-benefit analysis has been demonstrated or documented for LOSA to date, and the criterion of ‘resources’ as defined here does not consider benefit or value from investment, only that the use of resources is prohibitive. Future research could consider where resources might be reduced, or alternatively, could consider the benefits or savings from LOSA to justify the high upfront investment.

6.4.5 Acceptability of LOSA

Acceptability in the current framework refers to the degree to which all stakeholders can accept the underlying model of human performance and therefore the results generated by the method. For this criterion, LOSA scored highly (3). On the one hand, the endorsement and claimed benefits of LOSA outlined above and in Chapter 5 would seem to suggest high acceptability of both the TEM model and the outcomes from implementing LOSA. However, the TEM construct and LOSA observation method have also been the subject of criticism from theorists (Dekker and (2001), Hollnagel and Amalberti (2001) and Dekker (2003), for example). Acceptability is, therefore, best described as moderate to reflect this polarisation of views. In summary, the acceptability of LOSA and TEM is a strength of LOSA that should be capitalised on to gain acceptance for new methods in ground safety. However, when considering the acceptability of the results in terms of their validity, further work is necessary to determine whether the results of LOSA accurately reflect what they claim to measure.

6.4.6 Documentability of LOSA

According to the criteria developed above, LOSA would score a 3 in terms how well it is documented both as a method and for the purposes of periodic assessment. Guidelines for the implementation have been published by industry bodies such as ICAO (2002) and the FAA (2006). LOSA was developed as an audit tool to drive continuous improvement. The results are documented in terms of management reports and can also be stored within a central database that can be maintained by the airline or by external providers such as the LOSA collaborative. ICAO recommends repeating a LOSA audit every 3–5 years. The data can be used as a baseline measure against which to measure future performance or even as a comparison with other similar (anonymous) airlines within the LOSA collaborative database. Documentability and auditability are considered strengths of the current method that should be retained when adapting the method for ground safety.

6.4.7 Theoretical validity of LOSA

The validity criterion considers both the validity of the underlying theory or construct of human performance on which the tool is based, as well as the validity of the data collection methodology employed. The TEM model has been widely employed in a number of different applications, as discussed in Chapter 5. However, the TEM model originally proposed by the LOSA team at the university has neither been validated nor disproved. There is no evidence to confirm or refute that threats, errors, actions and outcomes actually occur in the way the original TEM model suggests. Further research is needed to investigate the model and the relationships proposed.

Validity here is also a function of the data collection and coding methodology employed. There are several well-structured aspects of the methodology that help to ensure the data collected is valid, such as large sample sizes, observer training, processes to preserve naturalistic behaviours, and data verification through consensus. There are also several aspects to the approach which may threaten its validity, such as the coding taxonomy discussed above and asking observers to code phenomena that are not directly observable, such as internal error mechanisms or links between threats, errors and undesired states. This process may become invalid if observers are not able to observe the cause and effects together, or if they instead rely on ‘backwards causality’, to infer links between phenomena, as discussed in Chapter 5.

There are several mechanisms by which validity could be improved, both in terms of the underlying TEM model and the coding process. One way to increase construct validity may be to stay at the level of what is observed, and restrict observers to the recording of the threats present and the behaviours of the crew, without implying that they are related. A simplified TEM model, based on what is observable, would further support construct validity. The model could simply describe the observable phenomenon and use the data from observations to statistically analyse associations between threats, errors and undesired states to determine which are more likely to coincide. Further clarification of the taxonomy into discrete categories would also help to improve the reliability of coding. Replacing error terms with neutral descriptors, such as ‘variation’ instead of ‘error’, may discourage observers from making assumptions about cause through the labelling of ‘error’.

There are, however, a number of processes within the methodology that are advantageous and should be retained, such as those described in the LOSA operating characteristics. Furthermore, increasing training and ensuring observers reach a certain standard of coding consistency may help to improve inter-rater reliability.

6.4.8 Summary of LOSA evaluation

Based on the discussion above, Table 2 summarises the scores presented for LOSA against the criteria described in the evaluation framework. The scores suggest that, although LOSA has been well received by industry, there are areas where improvements can be made to strengthen the methodology before adapting it for application in ground handling. These areas for improvement in LOSA help to define the development needs for new tool.

Table 2: Summary of LOSA evaluation against framework criteria

Evaluation Framework Criteria	LOSA evaluation scores
1.Comprehensiveness	2
2.Consistency	2
3.Usefulness	2
4.Use of Resources	1
5.Acceptability	3
6.Documentability	3
7.Validity	2

6.5 Summarising Development Needs for a New Tool

The evaluation criteria described and applied to LOSA in Table 2 above have helped to identify some of the strengths and opportunities for improving LOSA. These can be summarised as development needs for adapting a new method for ground safety:

6.5.1 Features of LOSA to be maintained

- Perceived usefulness, credibility and face validity of LOSA is a key strength that should be retained and built upon to retain the perceived usefulness of LOSA's benefits for industry.
- The new tools should build upon the LOSA program's structured processes to maintain consistency of the approach. This can be achieved through retaining the LOSA operating characteristics that encourage natural behaviour, such as large sample sizes, guaranteeing confidentiality, secure data repository, data protection, protocols for interventions and use of data etc.
- Documentability is a strength of the current method that should be retained to record a baseline, to assist repeatable periodic audits and drive continuous improvement.

6.5.2 Features of LOSA to modify or improve

- The development of a tool for grounds operations must be capable of describing the full range of possible human behaviours in ground operations. A comprehensive range of domain-specific codes is needed for describing ground crew performance.
- The observation method and coding should be restricted to observable phenomena and use.
- The TEM model should be simplified, should use more neutral terms and should ensure that codes are mutually exclusive, thereby avoiding ambiguity.

- The influence of threats and errors, and associations between them, should be investigated through post hoc statistical analyses rather than assumed by observers. Data from observations can be used to test and refine the TEM model.
- Ongoing monitoring and calibration of observers is needed to ensure consistency and reliability of data over time.

The next chapter will outline how the development needs identified in previous chapters as well as those described above can be been incorporated into the adaptation of a new method of normal operations monitoring for ground safety. Once a new method has been developed and tested, it will then be possible to apply the same evaluation framework from the table above, to evaluate how well the new method has met its development objectives.

Chapter 7: Development of a Normal Operations Monitoring Tool

Chapter 6 summarised the features of LOSA that should be retained and modified, and made recommendations for the development of a new tool that builds on the strengths of TEM and LOSA whilst addressing some of its weaknesses. This chapter proposes a new methodology known as Normal Operations Monitoring (NOM). The chapter describes the development of the NOM methodology, which attempts to retain the strengths of LOSA whilst incorporating the recommendations for improvement. The next chapter describes the implementation of NOM in a real ground handling operation.

7.1 Developing a New Tool

This section describes the modifications undertaken to develop the Normal Operations Monitoring model. The model and tools are based on LOSA and the TEM but also incorporate elements of other influential models discussed in Chapter 3. The resulting model and methodology are described in more detail below.

7.1.1 The case for Normal Operations Monitoring

The name given to LOSA, the 'Line Operation Safety Audit', made sense as at the time, as its purpose was to audit the transfer of CRM training to line operations in flight crew. Over time, the value of LOSA evolved from a training evaluation tool to a unique measurement of both safety and, specifically, human performance during normal operations. It is argued here that the real value of observation methodologies like LOSA is their ability to provide an indicator of human and safety performance in normal operations. To differentiate the new tool from flight operations LOSA, and to emphasis its primary objective, the new tool was named Normal Observation Monitoring, or NOM. Although in the current context NOM has been developed for ramp operations, the name, model and tools are not domain specific. NOM has been developed to be generic so that it could potentially be applied to any environment where more data is needed about human and safety performance in normal operations.

Whilst most high-hazard industries are aware that poor human performance contributes to somewhere between 70 and 90 per cent of all accidents (Shapell and Wiegmann 2004; Goos 2011; Peters 2014), few organisations routinely measure the standard of human performance in their operations. Safety departments traditionally measure many other indicators (such as LTIs, near misses, and hazards in the workplace), but tend to collect data about human error only by exception, through incident reports. An organisation's ability to routinely measure how humans are performing in the system, however, could offer new opportunities to influencing safety outcomes.

Monitoring normal operations provides a leading indicator of human performance but also of system performance, in terms of the output or product the system produces. It is perhaps the most direct measure of how a system is performing (in practice) rather than as imagined. Monitoring what actually happens helps an organisation to understand the differences between 'work-as-imagined by those who write procedures, and work-as-done—or actually enacted' (Braithwaite, Wears and Hollnagel 2016).

Understanding the current system performance, and how this differs from the desired performance, allows organisations to target interventions to address weaknesses and maximise strengths. Ongoing monitoring allows organisations to set system performance targets and monitor the effectiveness of efforts to drive continuous improvements. Naming the new tool 'Normal Operations Monitoring' emphasises this important focus and clearly distinguishes the tool from LOSA and its adaptive forms (such as R-LOSA).

Hollnagel Woods and Levenson (2007) suggested that the challenge for organisations was to measure the gap between 'the system as designed or imagined and the system as actually operated, to make the gap visible, and provide and provide a basis for learning and adaption where necessary' (p. 98). By measuring normal operations, we can develop an understanding of current system performance and where this differs from the work described in the procedures.

Aside from a renewed focus on normal operations, Chapter 6 identified a number of additional modifications that would be beneficial to address the methodological concerns raised in Chapter 5. These included the need to update and the original TEM model and simplify the terminology, through the development of domain specific codes, changes to the observation methodology and the need to investigate the relationships between elements of the new model. A detailed description of how these changes can be incorporated into the NOM method is provided below.

7.1.2 Adapting the TEM model and terminology

Chapter 6 highlighted the need to revise the TEM model to improve its construct validity. Several problems have been identified with the original model, such as the relationship between threats and errors and their consequences. Ma and Pedigo et al. (2011, p.2.), for example, note that the TEM diagram in the ICAO manual is outdated because the hierarchical stick-and-boxes diagram implies that every error has a threat. They go on to report through discussions with LOSA developer Klinect (2009) that this was found ‘not to be true through actual LOSA observations’, as most errors are ‘spontaneous errors’ without any previous threat (quotation from J. Klinect, reported in Ma et al. 2011). This suggests it is now appropriate that the model should be amended to allow for errors to occur in the absence of threats.

Similarly, difficulties have arisen with identifying the relationships between threats, errors and their consequences in the model, as discussed in Chapter 5. Links between elements in the model, for example, are not always observable. Whilst one error may lead to another, this is difficult to distinguish from two unrelated errors occurring in a similar time frame. This suggests the model should be reviewed to avoid the need for observers to make inferences about causation. Instead, it is proposed that the observer be restricted to coding only what is observable, such as the threats present, actions taken and the outcome occurring during the period of observation. Associations between phenomena observed can be assessed statistically, using the data collected from many observations. Improvements to data analysis are discussed later in this chapter.

The original LOSA terms have been shown to be problematic, as discussed in detail in chapters 5 and 6. New terms and definitions are therefore needed to overcome these issues. The table below proposes modifications to reduce subjectivity when assigning codes, through the use of more neutral descriptors. Whilst the LOSA definitions were flight operations-specific, NOM definitions could apply generically to any domain.

In Table 3 below, the two left-hand columns describes the original ICAO and FAA LOSA definitions, the third column proposes a modification and the fourth column provides the justification for the proposed change.

Table 3: Proposed modifications to TEM definitions for Normal Operations Monitoring.

Definitions from ICAO doc 9803 2002	Definitions from FAA AC 120-90 2006	Modified NOM definition	Justification
		Threats	
Threat: an external situation that must be managed by the cockpit crew during normal, everyday flights. Such events increase the operational complexity.	**Threat:** an event or error that occurs outside the influence of the flight crew (i.e. it was not caused by the flight crew), that increases the operational complexity of a flight, and requires crew attention and management if safety margins are to be maintained. (Note: the actions and errors of other parties are recorded as threats if they impact the task of the flight crew).	**Threat:** a situation or condition that occurs outside the influence of those being observed (i.e. it was not caused by the crew being observed), increases the operational complexity, and **may** require attention. It has the **potential** to increase risk to the operation. (Note: The actions or variations of other parties are recorded as threats if they impact the task of the crew being observed).	The proposed definition remains close to the original but removes reference to flight crew and does not assume all threats have to be managed to avoid a UOS or accident

Definitions from ICAO doc 9803 2002 (cont.)	Definitions from FAA AC 120-90 2006 (cont.)	Modified NOM definition (cont.)	Justification (cont.)
Errors/Variation			
Error: an action or inaction that leads to deviations from organisational or flight crew intentions or expectations. Errors in the operational context tend to reduce the margin of safety and increase the probability of accidents or incidents. Error Categories: • **Procedural error:** Deviation in the execution of regulations and/or operator procedures. • **Communication error:** miscommunication, misinterpretation, or failure to communicate etc. • **Proficiency error**: Lack of knowledge or psychomotor ('stick and rudder') skills. • **Operational decision error**: Decision-making error that is not standardized by regulations or operator.	**Error:** an action or inaction that leads to a deviation from crew or organisational intentions or expectations. Errors in the operational context tend to reduce the margin of safety and increase the probability of adverse events. Error Categories: • **Proficiency errors:** aircraft handling issues. • **Communication errors:** when information is incorrectly transmitted or interpreted. • **Operational decision errors:** which increase risk.	**Variation:** an action or inaction by the team under observation, which varies from the work as. An objective standard is needed to assess whether a variation has occurred by referring to standard such as a standard operating procedures or operations manuals. (The LOSA error categories are replaced by domain specific variation codes).	A variation defines a departure or variation from an agreed standard of the work 'as planned' It does not allude to causation or assume that a variation will reduce safety margins or increase the likelihood of adverse events. The term is intended to be value neutral and observable without interpretation based on an objective standard (such as the current rules). Identifying variations is intended to identify gaps between the work as planned and work as performed.
Intentional non-compliance error: Wilful deviation from regulations and/or operator procedures.	**Intentional violations:** deliberate departures from rules or procedures – a non-compliance with rules or procedures where the flight crew verbalise their plan to violate in the presence of the observer.	No equivalent.	All variations from procedures are recorded as variations. As the intention of the subject is not observable it is not coded as intentional or unintentional.

Definitions from ICAO doc 9803 2002 (cont.)	Definitions from FAA AC 120-90 2006 (cont.)	Modified NOM definition (cont.)	Justification (cont.)
Management actions			
Fail to respond: the lack of a flight crew response to an error because it was either ignored or undetected.	**Detected/no response:** both options are provided on the LOSA worksheet. No definition provided.	**Unmanaged**: no action was taken when it should have been, e.g. to manage a threat or variation.	'Detected' and 'Ignored' were both rejected as unobservable cognitive processes. Only action and inaction of the crew is observable.
Trap: an active flight crew response in which an error is detected and managed to an inconsequential outcome.	**Managed**: the threat, error or undesired aircraft state was resolved, leading to an inconsequential outcome.	**Managed**: action was taken to manage a threat, variation or UOS, which resulted in a positive outcome, in line with the organisation's expectations.	This code includes observable action and relates it to the outcome.
Exacerbate: a flight crew response in which an error is detected but the crew action or inaction allows it to induce an additional error, Undesired Aircraft State, incident or accident.	**Mismanaged:** an error that is linked to or induces additional error or an undesired aircraft state.	**Mismanaged:** action was taken but did not result in a positive outcome in line with the organisation's expectations. **Unmanaged:** no action was taken in response to the threat or variation	This code includes observable action and relates it to the outcome. Unmanaged was added to describe when no action was observed to be taken. (See below).
Outcomes			
Undesired Aircraft State (UAS): an outcome in which the aircraft is unnecessarily placed in a compromising situation that poses an increased risk to safety.	**Undesirable Aircraft State (UAS):** a position, condition, or attitude of an aircraft that clearly reduces safety margins and is a result of the flight crew's actions. It is a safety-compromising state that results from ineffective error management. A mismanaged UAS can induce an additional error, incident, or accident.	**Undesirable operational state (UOS):** an outcome that has the potential for consequential (unsafe outcomes) and is therefore defined as undesirable by the organisation.	'Aircraft' in UAS is changed to 'operational' (UOS) to be more generic. Undesirability is determined by the organisational expectations, however, the causes of the UOS are not assumed by the observer.

Definitions from ICAO doc 9803 2002 (cont.)	Definitions from FAA AC 120-90 2006 (cont.)	Modified NOM definition (cont.)	Justification (cont.)
Outcomes (cont.)			
Consequential/ Inconsequential: an outcome that indicates the alleviation of risk that was previously caused by an error. Alternatively, a consequential outcome can include an additional error, UAS or an accident.	**Consequential/ Inconsequential**: an inconsequential outcome indicates the alleviation of risk that was previously caused by an error. Alternatively, consequential outcomes include and additional error UAS or an incident or accident.	No equivalent	Observers are asked to note only the presence of threats, variations and outcomes, including incidents or accidents. However, they do not code links between threats, variations of their outcomes or infer their consequences. Associations between elements in the model are identified through post hoc statistical analysis.
Incident/Accident: any undesired ending that completes the activity sequence with a negative, terminal outcome.	**Incident/Accident**: reportable occurrences resulting from mismanaged undesired aircraft state (ICAO AC 2002).	**Incident/Accident**: any safety event required to be formally reported e.g. by the organisation's reporting procedures or national law.	The definition describes the outcome only and makes no assumption about causes.

The terms described for Normal Operations Monitoring differ from LOSA/TEM in an effort to achieve a number of objectives: firstly, they are generic, so they can apply to many domains outside the cockpit; secondly, they are restricted to observable phenomena – all terms that require interpretation have been removed. The main distinction from LOSA is the change from the use of the term 'error' to 'variation' (from a standard). A variation describes how and where work as performed varies from the work as planned. The term variation is neutral and does not imply whether the act itself is positive or negative for safety outcomes. It also negates the need to have different interpretations of error or its causes. If based on an agreed description of the work as planned, such as standard operating procedures or manuals of task descriptions, the term variation can become an objective standard that should improve the consistency and reliability of the coding process.

7.1.3 Proposing a new NOM model

Using the new terminology described above, a new NOM model is proposed for the collection of data. The proposed NOM model incorporates more recent theories of safety and accident causation (discussed in Chapter 3) that have emerged since the original TEM model was proposed in 2002. The NOM model also attempts to redefine the way in which threats and variations are related to each other. The internal boxes in the model represent the observable data that can be recorded during an observation: threats; variations; undesired operational states and their management, based on the definitions outlined in Table 3. Elements within the model shown in Figure 15: Proposed NOM model. are explored in more detail below.

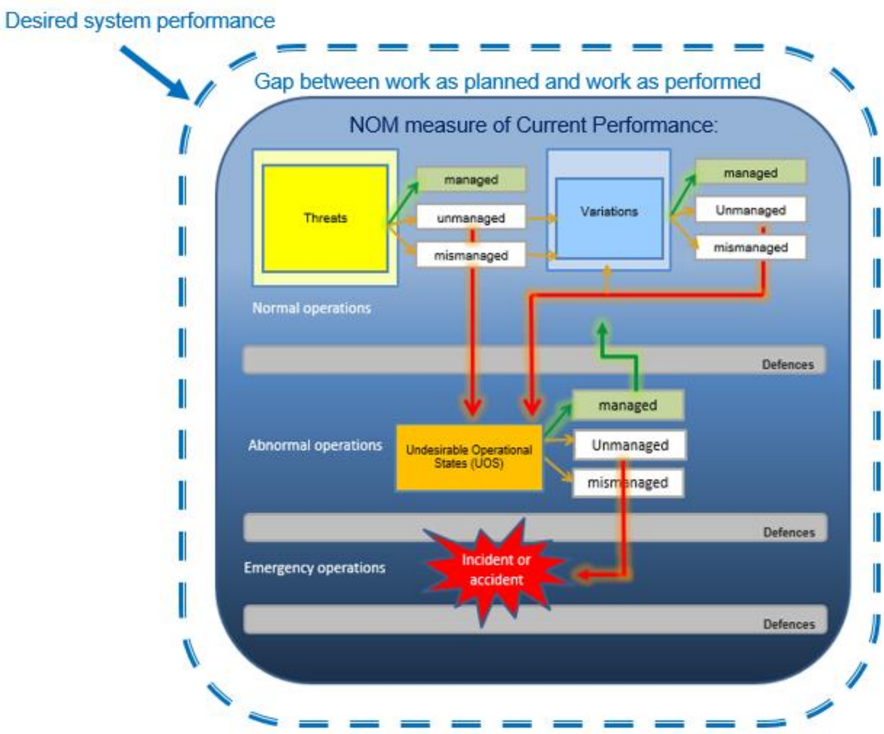


Figure 15: Proposed NOM model.

The NOM model attempts to position human behaviour within the broader organisational context. Unlike other accident models, such as Reason's, which depict organisations headed for failure (Walker and Bills 2008), the NOM recognises that normal operations are made of many every day actions which, most of the time, keep the operation operating safety. Often these behaviours will vary in some way from prescribed rules and procedures. Sometimes these variations create opportunities for undesired states to occur and allow accident trajectories to develop. These trajectories can also be prevented by the resilient management behaviours of people, as well as the organisation's defences. Inherent in the model is the assumption that, by collecting data about how tasks vary from prescribed plans, we may learn both how operations are normally kept safe and where accident pathways may develop. The objective is to identify strengths and weaknesses in current system performance in order to continuously evolve towards improved system performance.

Normal, abnormal and emergency operations: This NOM model recognises that most of the time threats and variations are managed to safe, inconsequential outcomes within normal operations. The model also identifies the ways in which an operation can shift from normal safe operating limits to abnormal or emergency modes. The model assumes that when a mismanaged threat or variation occurs, there is an opportunity for the operation to become abnormal, if an undesired operational state occurs. The red arrows in Figure 15 show where such pathways may develop and breach organisational defences. If the UOS is unmanaged or mismanaged, this creates a further opportunity for a consequential incident or accident to occur. Emergency operations are states where the harmful energy, as described by Haddon (1973), has already escaped. It is anticipated that such events would be extremely rare in NOM observations. Including the concept of emergency modes in the NOM model, however, may be useful to encourage organisations to think of all the ways in which energy escape or the defences available for minimising harm if incidents to occur (Haddon 1973; Kjellen 2000).

Threats, variations and Undesired Operational States: The NOM model shows that threats, variations and UOS can all be managed, unmanaged or mismanaged. The NOM model does not assume that all threats need to be managed to prevent variations from occurring. In a change from the way errors are depicted in the TEM model, in the NOM model variations can occur spontaneously, in the absence of threats. In LOSA observations, observers are asked to identify the links between threat and error management and their consequences. However, in NOM observations observers are asked only to record the presence or absence of threats, variations or Undesired Operational States and how they were managed.

Using this model, the observer is not asked to interpret the intention of the actors being observed or the nature of the internal mechanisms that contributed to their decisions and actions. Similarly, the observer is not asked to make assumptions about whether a threat or action is consequential. Observers only record the presence of threats, the actions of the subjects and note where these vary from the accepted standard. The observer also notes the operational outcome that follows including undesirable operational states (UOS), but is not required to link these to threats or variations. Relationships and associations between threats, variations and UOS are instead inferred afterwards using post hoc statistical analysis.

Accident trajectories and organisational defences: The red arrows in the NOM model are similar to the accident trajectories in the Reason Swiss Cheese model (1995). The red arrows depicted in the model indicate how an undesired state or an incident/accident could occur. The model assumes that breaking or disrupting this trajectory should limit these opportunities. Like the Reason model, the NOM recognises that the organisational defences (shown in grey horizontal shading) can stop an accident trajectory in its path. There are no codes developed for organisational defences; however, the observer is able to record any issues they identify at the organisational systems level. Including organisational defences in the model may also assist organisations to consider ways in which defences could be strengthened when considering interventions.

Management behaviours and resilience: The NOM model recognises that, most of the time, threats, variations and UOSs are managed within safe limits during normal operation. This model allows for both the variability and flexibility of people in the operation to respond to perturbations in the system by building on the Resilience concepts of Hollnagel, Woods and Leveson (2007), as discussed in Chapter 3. Responses that differ from SOPs are recorded as variations, but no assumptions are made about the intention of the behaviour. This recognises that some variations may be 'sensible adaptations' (Dekker 2002) and others may be 'intentional violations' (Helmreich 2000) as recorded in LOSA, but no assumptions are made in this regard during observations. The observer records the variation and how it was managed. The model also attempts to identify how crews' 'management' actions or resilient behaviours can keep an operation safe or return it to normal (safe) operations once a UOS has occurred; these are shown by the green arrows in the model (see Figure 15**Error! Reference source not found.**).

Measuring the gap between 'work as imagined' and 'work as done': Hollnagel (2011) describes 'work as imagined' as what system designers, managers, regulators and authorities believe happens or should happen, whereas 'work as done' describes what actually happens as individuals and teams adjust to the current operational conditions. Measuring variations from the work as prescribed through standard operating procedures or task manuals helps to illuminate the gap between policy and practice. By understanding this gap, it is possible to learn where the work as prescribed is impractical or in need or change, or where the system as designed may have not been fully thought through as to the real challenges of implementation in normal and abnormal operations. Catchpole and Jeffcott (2016) suggest direct observation as a way of identifying the difference between what is said and what is done as a basis for learning where change and adaption is needed. Identifying variations from prescribed tasks also identifies compliance issues where further investigation may be required into problems with training, competence, risk perception or culture.

7.1.4 Development of domain-specific codes

Chapter 6 identified the need for the new tool for ground operations will to be able to describe a complete range of human performance in the context of the operation being observed. To do this, a coding taxonomy will need to be developed to describe all the possible threats, variations and undesired states that could occur in the domain. MacDonald (2008) suggests that an independent system – derived from some criterion of adequacy, such as a rule or procedure – is needed to judge the appropriateness of any action. To develop a suitable classification system, it is proposed that the organisation's Standard Operating Procedures (SOPs) and manuals are used as the basis for this independent system. Variations are derived from all the ways behaviour might vary or deviate from this standard or 'criterion of adequacy' suggested by MacDonald.

In addition, a complete list of possible threats to the task will need to be identified. These should describe the performance shaping factors as outlined by Kirwan (1992). In line with the definition of threats above, this should describe a situation or condition that may need to be managed. A full range of possible threats and variations will need to be identified from sources such operating manuals, and task analysis, to describe the way tasks should normally be performed.

One mechanism for identifying the ways in which tasks may vary from the actions intended is to use a table of example guidewords. A 'Human-HAZOP' (Kletz) – based on a hazard and operability study – is a method that prompts consideration of all types of human failure. 'Human-HAZOP' guidewords are applied to each key step of the tasks to consider ways in which an operator could deviate. Variations can be identified by subjecting standard operating tasks and steps described in a task analysis to assessment. Application of the guidewords helps to identify all the ways in which they may be performed correctly. An example of possible guidewords for this purpose is provided in Table 4 below.

Table 4: HAZOP Guidewords (adapted from Kletz 1974).

Role: Loading Supervisor		**Task: Performing read-back to load control**	
Guidewords for identification of possible variations from standard operating procedure			
Action Variations	**Timing Variations:**	**Information/ communication variations**	**Checking variations**
No action	More time	More information	No check
More action	Less time	Less information	Part of check
Wrong action	Out of sequence	No information	Check wrong item
Part of action	Too late	Wrong information	Wrong check
Extra action	Too soon	Misinterpretation	Mistimed check

As well as variations and threats, the new classification system will also need to be capable of classifying all of the Undesirable Operational States (UOSs) that could occur. The potential unsafe situations will be particular to each domain, and even to each organisation. Lists of possible UOSs can be derived from accident and incident reporting, as well as from the regulations and their potential breaches, which result in undesired outcomes. The derived lists of possible threats and variations will need to be refined and verified through observations and through interviews or focus groups with subject matter experts to ensure the lists are comprehensive, necessary, and sufficient and fit for purpose for coding observation narratives. The end result would be a domain-specific set of codes based on the activity being observed.

7.1.5 Changes to the observation approach

Like LOSA, NOM collects data about safety and human performance in normal operations; however, whilst LOSA focuses on trained flight crews during normal flight operations, NOM will need to be adaptable to normal operations in other industries and domains. In each case, however, the specific scope of an observation will need to be clearly defined to ensure that data is consistently recorded. The scope should, for example, define all the tasks included (and excluded), as well as team members and personnel who will be observed, separately from the external parties that they may interact with.

Narratives require more effort, but are preferred to the checklist approach to recording activity during observations. This is so that the context, as it is experienced by the operators, is recorded for future verification purposes. Narratives written as prose provide the opportunity for qualitative analysis of the circumstances surrounding particular events. They can also be analysed for key words and grouped according to themes to offer another source of rich contextual information for further investigation of problems identified, which is not available through coding and checklists alone. The recording and coding of narratives should, however, be restricted to observable phenomena. Therefore, the NOM method would be unsuitable for highly cognitive tasks where the actions of operators are not visible.

Chapter 5 discussed methodological weaknesses of the LOSA method, which asks observers to make assumptions about cognitive processes such as ‘detecting’, ‘ignoring’, making ‘decision making errors’ or making ‘intentional violations’, all of which exceed the limits of an observation methodology. The new NOM model proposed above should assist observers in describing and coding observable phenomena together with the refined terms. Although the narratives can record as much contextual detail and activity as it is possible to record in the time available, the coding in NOM is restricted to observable elements in the NOM model. This includes threats and variations from standard operating procedures and the occurrence of undesired operational states. Each observer then codes whether these are managed (dealt with successfully), mismanaged (dealt with unsuccessfully) or unmanaged (not dealt with) based on the observable outcomes. Unlike LOSA, however, observers do not try to link threats and errors to undesired states during observations. Instead, these relationships maybe inferred from a statistical post hoc analysis.

7.1.6 Changes to the training and calibration of observers

Chapter 6 identified the need for observers to be subject matter experts, ideally selected from the population of personnel they are going to be observing. This would allow them to understand the operational perspective of those being observed and the important aspects of the context when recording narratives. Observers will also need to be extremely familiar with the standard operating procedures in order to identify variations. For this reason, it is suggested that only experienced personnel are recruited.

Chapter 6 also identified the need (as per Fletcher et al. 2013; Jepsen et al. 2015) for observers to be extensively trained in the observation methodology and codes. The training of observers should be focused on ensuring that observers can consistently and reliably code behaviours in the same way. Observers will need to be calibrated to a certain standard of accuracy when applying the codes before data collection can begin. Ongoing monitoring and calibration of observers is needed to ensure consistency and reliability of data over time. Processes need to be put in place to ensure that the tools and codes remain stable over time, so that the process can be repeated with reliable results.

7.1.7 Maintaining the original operational characteristics

Chapter 6 identified the features of LOSA that should be retained and suggested that the new tool should build upon the LOSA program's structured processes, such as those described in the LOSA operating characteristics (see Chapter 5). Adaptations have been made to the original characteristics described by Klinect, Murray, Merritt and Helmreich, (2003, p. 3) for application in the NOM process, so that they can be applied generally to any industry setting.

1. Observations should be restricted to normal operations – as with LOSA, NOM should avoid observing non-normal situations such as training exercises or combining NOM with other audits where people may not be behaving normally. Like LOSA, NOM observers should be trained not to interfere in a way that could influence normal operations.

2. Anonymous and confidential data collection – observations must be conducted under conditions that protect the anonymity of those being observed. LOSA protocols for data protection are also observed in NOM so that NOM observers do not record names or other potentially identifying information.
3. Voluntary participation – as with LOSA, personnel can refuse to participate in NOM observations. If the aims and data protection protocols of the program are well communicated, refusals should be minimal; if a high number occur, this indicates a low level of trust in the process, which may mean observations of normal behaviour may not be achievable.
4. Joint management and employee cooperation – like LOSA, the NOM program should have the support of both management and the employee population in order to ensure there is trust and confidence in the process to encourage normal behaviour.
5. Safety-targeted data collection form – NOM uses an observation form for collecting general demographic information such as the location, time of day or specific circumstances of the task being performed; however, these details should not be able to identify individuals in keeping with the original LOSA data protection protocols.
6. Trusted and trained (peer) observers – as with LOSA, NOM observers must be subject matter experts, chosen as trusted and experienced operators from the population being observed. They must be very familiar with the current rules and procedures, selected on the basis of their observational skills and extensively trained in both the theory and practice of recording and coding narratives.
7. Secure data repository – to assure confidentiality, NOM programs, like LOSA, must have a trusted data collection site or, alternatively, data can be collected and stored by an independent third party.
8. Data washing/verification – consistency checks are needed to assure quality data. After all the observations are complete and the data have been entered, a joint data cleaning roundtable should be convened to review all the recorded threats and variations. As with LOSA data washing, it is the job of the NOM roundtable to manage

discrepancies in coding. Where the wash team disagree with the original code allocated by observer, the wash team determine by consensus the code that should be allocated. This process can identify coding errors and help to provide consistency of coding where there are differences in interpretation by individual observers.

9. Data-derived targets for enhancement – statistical analysis of the data should investigate trends in the data such as high prevalence threats and variations but also the associations between elements of the NOM model. As with the LOSA analysis, this data should be used to identify targets for interventions that will have the most impact on safety. Periodic monitoring (repeated NOM) can be used to identify whether interventions were effective.
10. Communicating results through feedback – participants in the NOM program will want feedback on the results as well as the management activities and interventions planned to drive continuous improvement. As with LOSA, documenting the results as a baseline is important to assist repeatable periodic audits and drive continuous improvement.

As well as retaining the ten LOSA operating characteristics described above, Chapter 6 also identified other desirable features of the LOSA approach that should be retained, such as such as large sample sizes to ensure that the data is representative of normal operations generally. LOSA guidance recommends observing at least one per cent of the operation (FAA 2006). NOM should also therefore attempt to ensure that the sampling of tasks observed is random and broad enough to be representative of normal operations, so that results are generalisable.

7.1.8 Use of NOM for identifying targets for training

The original purpose of LOSA was to evaluate Crew Resource Management training and set targets for improvement. Whilst CRM training is commonplace for flight crew, it is rare outside the cockpit in aviation. Although attempts have been made to take flight crew CRM training topics and adapt the material to other industries, there is insufficient evidence to demonstrate the benefits of this approach (Salas et al. 2006). It is proposed, therefore, that NOM observations should not attempt to record how well a team performs against a set of behavioural targets designed for pilots, as there would be little benefit in using NOM for this purpose.

Instead, NOM should attempt to identify, with data from observations, which behaviours should become targets for future training. NOM observations can record resilient behaviours as well as when actions do not go as planned. Data from NOM observations can, therefore, highlight successful and unsuccessful behaviours based on the outcomes that result. Post hoc statistical analysis can be used to identify behaviours associated with desirable/ undesirable outcomes so that these can become the targets for future risk-based training interventions. The narratives can also supply realistic examples from operational settings to be developed as learning case studies. Future audits could measure training transfer once a training intervention has been implemented.

7.1.9 Post hoc statistical analysis of relationships

In NOM, the influence of threats and errors (and associations between them) should be investigated through post hoc statistical analysis rather than assumed by observers. Statistical tests can be used to identify significant trends in the data and the strength of relationships between different threats, variations and undesired states. The observations can also record a great deal of demographic information, which can be statistically analysed for trends.

LOSA observations, for example, record information such as the aircraft type and the city pairs being flown to and from. This information can be useful in identifying issues that occur across a particular fleet or route. Similarly, ground observations could collect general information about the task conditions, team size and time of day or equipment used, for example, to assess the impact of any of these features on performance outcomes.

Reporting of absolute numbers in LOSA reports has led to criticisms that LOSA focuses on counting negative events (Dekker 2007). Care should be taken to ensure the new method is able to provide objective data from observation but also to make use of qualitative contextual data so that variations are not reported out of context. Narratives provide the rich context necessary to understand responses that made sense at the time, and can help to analyse which behaviours were most adaptive in managing threats and which lead to further complications.

Complementary investigative methods will be needed to interpret findings from the data analysis and provide a diagnosis to inform the design of safety interventions. This may include a qualitative analysis of the narratives or interviews with subject matter experts. Hypotheses developed through qualitative investigation could then be further investigated. Access to a reference group of domain subject matter experts would, therefore, be very useful when designing a new methodology for ground operations. Finally, the data from statistical analysis should be used to test and further refine the new NOM model where possible.

7.2 Summary and conclusion

This chapter has provided a description of a new NOM model and observation model. It has attempted to build on the strengths of the original LOSA method whilst overcoming some of its limitations and methodological difficulties as identified in previous chapters. NOM has been developed to be generic in terms of its underlying model and approach; however, it is recognised that significant adaptation will be required to adapt the NOM tools to a particular domain. The next chapter describes the adaptation of NOM to the ground handling domain.

Chapter 8: Implementation of a Normal Operations Monitoring Tool in Ground Operations

This chapter describes the implementation of the NOM methodology within a ground handling environment, including the aims of the trial and method used. The first section provides a description of the operation and loading task to be studies with the NOM application. This is followed by a description of the aims of the study, the development of customised ramp specific codes and tools and finally the application of the NOM method in the live operation. The results of this NOM application are presented in chapters 9 and 10. Evaluations of the NOM method against the evaluation framework criteria proposed in Chapter 6 are discussed further in Chapter 11.

8.1 Aims of the Normal Operations Monitoring (NOM) Trial

The main aim of the trial was to customise and test the NOM method and tools in a live operational environment. The objective was to use the method to analyse a baseline measure of current performance and its variation from the work system as planned in the standard operating procedures. In addition, the aim of NOM is assess the impact of the threats and variations on safety outcomes. It was hoped that the NOM trial would provide novel data about human and safety performance on the ramp, which could inform safety interventions on the ramp. The trial also provided an opportunity to test the NOM model, tools and observation method and evaluate the NOM approach against the framework established in Chapter 6.

8.2 Description of the Ground Operations Test Site

8.2.1 Test site for NOM trial

The trial NOM implementation included ground handling organisations and airlines operating a range of aircraft types for both domestic and international flights over a 7-year period between 2008 and 2015.

The task under observation was the unloading and loading of the aircraft turnaround by ramp crew. Although a number of different teams and personnel access the aircraft during the turnaround, the focus of the observations would be on the ramp crew – which included the loading supervisor – and their loading team.

8.2.2 Test site Standard Operating Procedures

Chapter 7 described how the NOM Model codes variations from a standard. For the purposes of the current study, the objective standards used to describe how a task should be performed include the relevant ramp services manuals, and standard operating procedures (SOP's) for aircraft loading and operating equipment as well as the relevant airport's safety rules for airside driving. When the ramp crews' actions vary from the task described in the manual SOPs or airport regulations, they are recorded as variations (note that only the ramp team's activities can be coded for variations against a standard; the actions of personnel from other departments are coded as threats to the ramp crew).

Before describing the specifics of how the NOM method was adapted for this operational environment, it is helpful to first understand the task and activities to be observed and the types of issues and challenges that the crew must manage. A general summary of the loading task, from how the load is first planned, through to the aircraft being approved for take-off, is provided below.

8.2.3 Description of the task to be observed

The task to be observed included all of the airport ramp team's activities during the turnaround process, from before the aircraft's arrival at the bay until the aircraft departs the bay. A description of the steps and activities involved in the turnaround process is provided in more detail below.

Planning of the aircraft load. The SOP's describe the load controller responsibilities for planning the position and total weight of the load, which is used to calculate fuel ratios. The load sheet, or Loading Instruction Report (LIR), provides the ramp crew team leader – known as the Loading Supervisor (LS) – with the positions for each load to ensure the aircraft is evenly balanced or 'in trim'. The planning and balance of the load depends on the correct paperwork being issued by Load Control and distributed to the different team members of the ramp crew, to allow the load to be distributed exactly as planned on the LIR.

Planning and allocation of ramp crew to aircraft. Ramp crew teams are allocated to aircraft turnarounds by a range of different rostering systems that assign available personnel to arriving aircraft. The crew size allocated to each aircraft will depend on the size of the aircraft, but also on the amount of freight and cargo to be handled. When rostering issues occur they can cause difficulties, such as split teams, team shortages and ramp team members performing multiple roles.

Preparation of the aircraft bay. The SOPs describe the ramp crews' responsibilities for preparing the bay for the aircraft arrival, by ensuring that all equipment is in place and staged safely behind equipment lines. This can be challenging if there is not enough space available or the staging areas are not clearly defined. Equipment and vehicles fouling the bay or access paths can also contribute to congestion on the ramp. Bay congestion can be exacerbated by the design of the terminal itself, as well as the incorrect positioning of equipment and vehicles by ramp crews and other personnel who access the ramp. Preparing the bay also involves ensuring that the bay is clear of Foreign Object Debris (FOD) that could damage the aircraft. FOD refers to any material that fouls the airport ramp, which can be ingested by aircraft engines and cause damage. The most typical sources FOD are plastic and cardboard, metal locks from baggage and plastic straps used on mail boxes. FOD bins are provided at various locations around airports for depositing FOD material.

Approaching the aircraft. SOPs describe how and when the ramp crew can approach an aircraft. Crew must wait for the anti-collision beacon to be turned off and for the engineer to give the signal that it is safe to approach the aircraft by giving a 'thumps up'. SOPs state that, prior to opening the cargo doors, the crew should inspect the doors and sills for any damage that may occurred prior to arrival. Most ground handlers and airlines have specific rules for driving up to and around aircraft; for example, conducting a brake test 5 metres from the aircraft on approach and driving within walking pace when within 2 metres of an aircraft. This is referred to as the 'circle of safety' procedure (IATA Airport Handling Manual 2010).

Unloading the aircraft. Each member of the ramp team is responsible for obtaining a copy of the Offload Instruction Report (OIR) and the Loading Instruction Report (LIR). The Loading Supervisor is responsible for ensuring that the crew have the latest documentation. If there is a mismatch between the documentation and the load arriving, this is recorded as an irregular load. An irregular load is any occasion where the actual load on board does not reflect the planned/documented load. Arriving irregular loads present a threat to be managed by the ramp crew under observation, attributed to a problem originating at the previous port of departure.

Use of Ground Servicing Equipment. The crew use a range of Ground Servicing Equipment (GSE) to move cargo to and from the aircraft. Belt loaders have conveyor belts that transport cargo from the ground to the aircraft hold. The SOPs describe how the side rails, sometime referred to as 'batwings', must be in the lowered position when in motion to prevent damage to aircraft (especially if they are up) when driven under the aircraft wings or fuselage. There are different types of ground servicing equipment used for different types of aircraft. Using the wrong equipment can also cause damage to the aircraft and misuse of equipment can injure personnel.

Loading the aircraft. SOPs describe the Loading Supervisor's (LS) responsibilities for ensuring the safe unloading and loading of the aircraft according to the weight and balance Load Instruction Report (LIR) prepared by Load Control. The Load Instruction Report (LIR), or load sheet, describes the cargo freight and bags to be loaded into each position on the aircraft, including assigned positions for livestock and dangerous goods. Aircraft loading checklists are distributed to the hold operators to check off loads against their assigned position. However, it is the Loading Supervisor's responsibility to provide the required documentation to state that the aircraft has been correctly loaded. Failure to load the aircraft according to the LIR, or failure to provide the completed documentation, can lead to regulatory breaches and fines.

Managing special loads (e.g. livestock/ dangerous goods). The regulations for the transport of special cargo, such as live animals and dangerous goods, describe the handling requirements that must be in place. Where there are included in the cargo, the LIR should provide 'special instructions' for such items. 'Dangerous goods' describes any materials that pose a risk to health, safety, property or the environment when transported by air, including explosives, radioactive materials, flammable liquids, dangerous or volatile chemicals, strong acids, compressed gases, poisons and aerosols. When domestic animals or livestock are to be loaded in an aircraft hold, they are given a transport code of 'AVI' and special instructions apply to their loading. For example, it is mandatory for animal cages to be restrained by company-approved tie-down restraints to prevent movement during flight, and there are procedures for ensuring adequate ventilation around the cage. If any freight or cargo is loaded in an unsafe manner, e.g. without the correct restraints or protection, when transporting dangerous goods, or is not in the position it was assigned by Load Control, this is an irregular load and also a regulatory breach

Securing the load. SOPs describe how the load, and load positions, within the hold must be secured before take-off. Position locks, webbing and restraints are used to ensure that loads do not shift during transit. Even empty positions are required to have raised locks to ensure cargo cannot move into empty positions. The Loading Supervisor is required to visually inspect the restraints and ensure that door sills, stops lateral guides webbing and stanchions are in place. Once the load is secured and signed off by the Loading Supervisor, the hold door cannot be reopened without permission from Load Control and the engineer.

Performing the read back. Once the load is secure, the SOPs describe the Loading Supervisor's responsibilities for contacting Load Control and performing a verbal confirmation or read back. The read back is the protocol for confirming to Load Control that the load is secure prior to take off. Providing the read back before the load is on board, or failing to provide the required information during the read back, are all considered to be regulatory breaches.

This section has provided a general description of the organisation and tasks where the NOM method was applied. The next section provides a description of the NOM study and its application at this site.

8.3 Development of Customised Ramp-Specific Coding Tools for the Trial

The NOM tools were based on the new NOM model and codes defined in Chapter 7. Codes were developed to describe all the possible threats, variations that could be encountered during an observation including actions taken, breaches of defences and accidents . To come up with as complete a list as possible, a number of activities were undertaken to develop threats and variation codes. The process involved developing a list of tasks based on preliminary observations of ramp activities, a review of manuals and standard operating procedures, risks and controls and the development of a task analysis to describe the activities of ramp crew. This task analysis was then assessed for all the ways that a task could be performed incorrectly, thereby developing a list of variations. To refine and embellish the taxonomy of codes, interviews were conducted with ramp crew and workshops with subject matter experts (SMEs) until a final taxonomy of codes was produced for use in observations. These activities were undertaken in four steps, described in more detail below:

8.3.1 Identifying tasks and how they should be performed:

Preliminary observations. Observations of ramp activity were made unobtrusively from the public area of an airport terminal with the researcher and a Subject Matter Expert (SME). Observable threats, variations, actions, outcomes and defences or controls were recorded in a list for further classification and refinement.

Review of manuals and Standard Operating Procedures. Several Ramp manuals, current regulations and standard operating procedures (SOPs) from different airlines were reviewed to identify at a high level how loading tasks were supposed to be performed. This helped to define a common standard from which the actions of ramp could vary, to provide a list of potential variations. SOP's naturally varied between organisations, and even geographical locations within the same organisation. In such cases generic versions of codes were sought that captured the essences of the process being described.

Task analysis. Using the information from observation and the review of manuals and SOPs, a task analysis was developed to represent how the task should be performed according to the current standards. This helped to identify all the ways in which the task could vary from the current rules and procedures.

8.3.2 Identifying all the ways in which a task could vary from SOPs

HAZOP error analysis workshop. Subject Matter Experts from the ramp environment were presented with the task analysis and asked to consider all the ways in which the tasks described might unfold in unexpected ways, including possible variations from the standard operating procedures or regulations and the controls or decefences that might be in place to prevent or contain harmful outcomes. To guide this process, each task was assessed using guidewords from the human hazard and operability (HAZOP) table (Kletz 1974; Swann and Preston 1995; Kirwan and Ainsworth 1992). The HAZOP guidewords provided prompts for the consideration of all the ways in which the steps in a task might go wrong. This process produced a preliminary list of variation codes.

8.3.3 Identifying threats, variations undesired states and defences from risk assessments and incident data

Incident reports provided examples of conditions or circumstances that were influential in an incident. Ramp safety incident reports from participating organisaitons, and from the ATSB and FAA, were reviewed for examples of threats, variations, undesirable states, breaches of defences and accident scenarios.

CASA definitions of regulatory breaches (such as irregular loads) were also added to the list of undesirable operational states. These lists were further refined with reference to each organisations incident database descriptions of crew actions that lead to undesirable operational outcomes or regulatory breaches. The result was a list of potential codes for threats, variations and undesired operational states as well as categories of actions taken, defences and controls and potential accident scenarios.

Interviews. Ramp crew members were asked to describe situations or conditions that made their tasks more difficult or that increased risk to the operation. They were also asked to describe things that could go wrong when activity varied from what was intended or planned and the controls or defences that were meant to be in place to prevent incidents from occurring. The data provided from the interviews was used to embellish and further refine a list of codes suitable for the loading task.

8.3.4 Refining the coding taxonomy

Workshop with ramp personnel. A workshop was held to further develop a taxonomy of necessary and sufficient codes for use in ramp observations from the preliminary list of codes. The aim was to describe the full range of activity that might be observed during a ramp observation, while avoiding duplication within the coding set. All the potential codes in the initial list were sorted into groups or 'major categories' and presented to a workshop of SMEs from the ramp for review. Any items were checked to ensure they met the criteria for each definition in the NOM model (described in Chapter 7, Table 3). The categories were also reviewed for any duplication or overlapping codes. The result was taxonomy of ramp observation codes categorised into major categories and subcategories.

Following this review, codes were aggregated into 13 major threat categories, 11 major variation categories and 9 major Undesirable Operational State categories. Some examples of the codes developed for NOM observations are outlined in Table 5 below.

Once the coding taxonomy was developed, the database for the recoding, coding and analysing of NOM data was designed to manage the coding of narratives. A priority for the design was speed of data entry and automating as much of the functionality as possible, in order to reduce the coding time and resources required.

The database has one demographic data entry screen; however, no individuals can be identified using the software. There is a separate tab for the observer to enter their narrative of each turnaround. The observer's handwritten narrative notes are then typed up into free text fields in the NOM software.

Once tools were developed, they were tested for use with real turnarounds at an international airport by two observers. The layout and functionality of the database was changed to reduce data entry errors and to speed up the coding process through improved drop-down menus.

8.4 Selection and Training of Observers

8.4.1 Selection of observers

Observers were selected by the participating organisations based on selection criteria provide below. They were not paid any additional funds for their participation except where shift penalties applied under their individual awards. The following recruitment criteria, suggested by the LOSA collaborative (ICAO AC), were used to select observers:

- computer literacy,
- knowledge of Standard Operating Procedures and operation,
- credibility amongst peers,
- loading Supervisor qualification,
- trustworthiness/ability to maintain confidentiality,
- safety focus,
- highly developed interpersonal skills, written and verbal,
- attention to detail, and
- experience with working in a team environment.

A set of competencies for observers was developed based on these characteristics. The selection process included a practical test of narrative writing and minimum qualifications at the Loading Supervisor level or equivalent were required. All remaining candidates were interviewed using a set of five standard questions. Following the interview, the observers were given briefings on how to write a narrative, and then asked to watch a staged video of a ramp turnaround that included a number of threats and variations. The narratives were then graded based on quality, detail and length, as well as ability to pick up on relevant threats and variations. In total 13 observers were selected for training as observers.

8.4.2 Training of Observers

The selected observers were given approximately five days of training. This included a 3-day training-course covering theory and conceptual models, an overview of the NOM methodology, practical skills for narrative writing and coding, as well as practical experience using the software. Videos of ramp activity were used for practice observations where available. Classroom training was followed by two days of observation training on the ramp with feedback. An summary of information covered in the training modules is provided in Table 6

Table 5: Observer training modules and content

Training Module	Training Module Content
Module 1	Introduction – Aims and objective of the study and the role of the observer.
Module 2	History of LOSA and TEM and more recent developments in contemporary safety science
Module 3	Introduction to NOM and tools.
Module 4	Identifying threats, variations and Undesired Operational States, management actions, defences and accident types.
Module 5	Correctly applying the codes.
Module 6	How to conduct an observation on the ramp – briefing, and conduct.
Module 7	Narrative writing
Module 8	Observer calibration exercises
Module 9	Reviewing results and providing feedback in order to standardise observer coding.
Module 10	Data entry and software training.
Module 11	The data verification process (data 'wash').
Module 12	Summary, including how observation data will be used.

During the course, the observers are trained to act as a 'fly on the wall' to avoid interfering with the natural behaviours of the ramp crew as much as possible. Observers were trained only to intervene if they perceived an immediate and serious risk to personnel or aircraft during the turnaround.

At the end of the classroom training, the observers were invited to conduct practice observations at an airport in pairs. Observers were encouraged to code practice observations independently and then compare the consistency of coding with their partner. When paired observers had established coding consistency with each other they returned to the class room to compare their results with the rest of the observer group. The group discussed differences in coding where they occurred, so that consensus between observers could be reached. The training video was revisited to test the degree of coding consistency between observers. When all the members of the group reached agreement and were able to code the video in the same way, the training was concluded.

8.5 Communicating the Aims of the Trial

The preparation for implementing the NOM program followed the guidance provided by ICAO and the FAA for LOSA implementation (described in Chapter 6) to preserve the confidentiality of the observations and build confidence with ramp crews about the purpose of the program. To support the implementation of NOM, a brochure was developed for observers to hand out at participating ports, to communicate the confidential nature of observations.

8.6 Testing of Tools and Observation Methodology

The implementation of the method was conducted in a number of test phases to allow for incremental improvements. The objective was to test the suitability of these ramp-specific tools and processes, including the coding system, software and database, through a small number of observations. The aim was to gather feedback from the test phase observer on the practical usage of the observer forms and observation approach in general. Based on the feedback provided a number of improvements were made to the tools, including:

- reducing and regrouping codes to improve the drop-down menus;
- improved navigation tools within the database
- a mechanism for flagging when no suitable code exists and a process for adding additional codes
- automated error notifications for conflicting data or blank required fields
- The need to include observers from a range of different organisations n the data verification process (as described in Chapter 7) to help manage and interpret differences between each organisation's SOPs.

8.7 Procedure Used for Implementing the International NOM Trial

8.7.1 Subjects

The NOM observations included all ramp activity within the turnaround described in section 8.1. The NOM observations focused on the actions and interactions of the ramp loading teams. These teams usually consisted of a Team Leader (or Loading Supervisor) and between four and nine team members from the ground service provider, which could include labour supply contractors.

8.7.2 Sampling method

The scope of the operation to be observed included operations across 28 airports internationally. The observers travelled in pairs to each port location. In larger ports, observations were scheduled at predetermined times to give better coverage of night shifts, day shifts and during peak periods. However, in smaller ports observations took place whenever aircraft were available. Flight arrivals are not evenly distributed across different times of day and, in some ports, night flights are restricted by curfews, reducing the opportunity to conduct night observations. Over 1374 observations were conducted over a 7-year period; however, due to some data entry errors and incomplete observations, only 1302 complete observations were included in the analysis. Observers were randomly assigned to incoming flights based on aircraft arrivals as they came up on each airport's scheduling technology.

8.7.3 Apparatuses

Each observer was provided with the following apparatuses for data-collection:

- The observation forms and notebooks
- The NOM software, database and laptops
- Information packs (brochures and supporting information about the NOM program to communicate to ramp crew, prior to beginning an observation)

8.7.4 Conducting observations

Before beginning any observations, a briefing was held explaining the aims of the program and the non-jeopardy, confidential nature of the data collection process. At the start of each observation, the observer introduced themselves to the crew and handed out a brochure about the observation process, explaining how data would be used. Subjects had an opportunity to ask questions and to refuse to take part in the observation. In foreign ports, the observers were introduced by their security escort in the local language. If there were no objections, the observer explained how the observation would be conducted and reiterated the confidentiality of the process.

The observers then recorded demographic information about the flight and load. When each turnaround began, the observer recorded a free-text narrative of the activities as they unfolded. It was recognised that the observer could not record all the activity that occurred during a turnaround. Observers were instructed to 'follow the action' – that is, to go wherever the focus of activity was at any one time, but, when in doubt, to shadow the Loading Supervisor.

Narratives were handwritten to allow for a flowing sequence of events recorded in real time during the turnaround. This included recording the interactions of the ramp team with other airport personnel involved in catering or freight and baggage, load control or flight operations etc. The observer recorded any threats that were present, the actions taken by the crew and any variations from standard or expected practices. The observer noted the crews' response to threats, variations and undesired operational states and the actions taken to manage these events. The observers described the safety outcomes of the activity observed and the effectiveness of any controls or defences where observed. Finally, the observer thanked the crew concluded the observation.

8.7.5 Coding

Assigning codes. After completing the observation, the narratives were typed up and entered in to the NOM database. Drop-down menus, containing the list of grouped codes for major categories and sub-categories, were displayed at the side of the free-text narrative field. Observers highlighted and 'tagged' text in the narrative to be coded and assigned a code number using the drop-down menus . Prompts within the database helped to ensure that all the necessary fields were completed before the observation was submitted for analysis.

8.7.6 Data-verification ('data-washing')

Data verification teams were assigned from the observer pool. Each group reviewed the coded observations to ensure the coding assigned was appropriate based on the information in the narrative. Where disputes arose, coding was assigned via consensus of the wash team. Where there was agreement that a code had been wrongly assigned, the wash team was able to alter the original code. During the data-wash process, the wash team also noted issues of interest such as any recurring themes or questions that could be further investigated during the analysis stage.

8.7.7. Data analysis

Data analysis. The data was assessed using a statistical software package (SPSS). A summary of the general results and the statistical analysis are provided in chapters 9 and 10. The issues and comments fields completed by observers were downloaded from the database and filtered by port, as well as being grouped into themes across the ports for further analysis. For example, if the issues raised questions requiring clarification, such as the interpretation of rules or local conditions, these issues were followed up with the relevant organisations for clarification.

Follow-up investigation. Follow-up investigative methods included interrogation of the observation narratives, including keyword searches, and interviews with subject matter experts to try and understand more about the causes of issues observed and to inform intervention strategies. Over 20 interviews were conducted with ramp personnel. During the interviews, ramp crew members were asked to provide additional detail and interpretations about why certain behaviours might be occurring .The semi-structured interviews provided additional detail and lines of inquiry to assist the analysis and interpret the data.

8.7.8 Observer monitoring and calibration analysis

A process for analysing differences between observers was introduced to monitor the consistency of coding across the observer group. This involved developing a profile of each observer's coding tendencies and plotting them on a graph. Observers were compared in terms of the rate of threats and variations they recorded per observation, as well as their standard deviation from the average rate per observation from all observers. An example of this is shown in Figure 16, where Observer 3 had a higher rate of variations per observation than other observers.

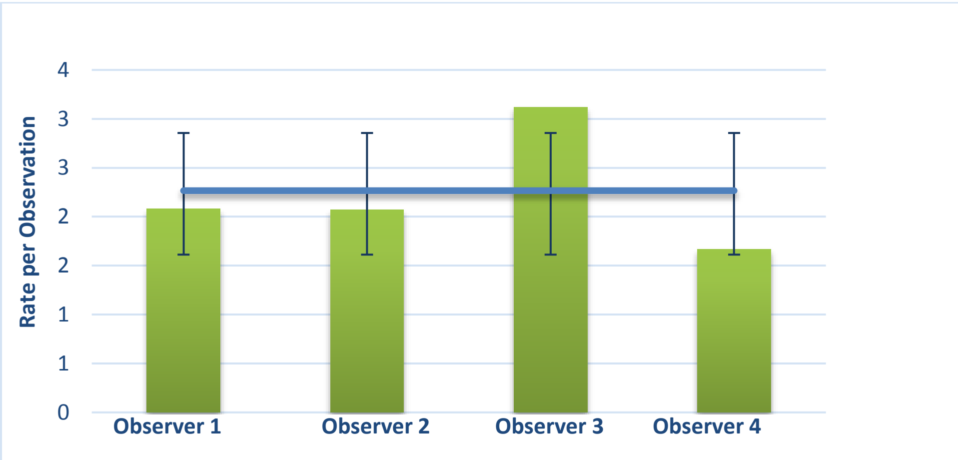


Figure 16: Observer threat rate compared against the average

As part of an ongoing feedback and coaching process, these individual profiles were shared with each observer so they could see how their coding compared against the average. Where outliers were identified – for example, where one observer records a particular code much more frequently than the average – these differences were discussed with the observer group. Feedback and discussions with observers helped to support the ongoing calibration of observers in their use of codes. The results of the data collected from this NOM implementation are presented in chapters 9 and 10 and assessed against the evaluation framework criteria in Chapter 11.

Chapter 9: Results of NOM Implementation in Ground Handling Operations

This chapter provides an overview of the data collected from 1302 NOM observations conducted within during normal operations in a ground handling environment. The observation forms collected general information about location, aircraft and crew involved in each observation. The observations themselves provided quantitative information from the coding and analysis of threats, variations and undesirable operational states (UOSs), whereas the narratives, observer issues analysis and port reports provided qualitative data about the specific conditions and context. This chapter describes the data collected from the observation forms and provides and from the coded observation narratives in six sections, as outlined below.

Section 9.1 (Analysis of observation forms). This section provides a general overview of the information collected from the observation forms themselves, including the location, aircraft type, crew and activity observed, time of day and whether the aircraft arrived early or late.

Section 9.2 (Analysis of threats). This section analyses data regarding threats the ramp crew are faced with. The major threat categories are analysed and broken down to show the contribution of each threat type or subcategory. The most frequently occurring threats are highlighted. Excerpts from narratives, port reports and observer issues provide examples and additional context for interpreting threat data.

Section 9.3 (Analysis of variations). This section analyses variations from standard operating procedures by ramp crew. The major variation categories are analysed and broken down into subcategories, to understand the contribution of each variation type. Variations are also analysed by port to indicate locations where rates were higher for some variations. The most frequently occurring variations are highlighted. Excerpts from narratives, port reports and observer issues provide examples and additional context for interpreting variation data.

Section 9.4 (Analysis of Undesired Operational States). This section provides an analysis of the Undesired Operational States, or UOSs, recorded during the observation period. The major categories of UOSs are presented and broken down to show the contribution of each UOS type. Qualitative examples of each UOS are provided from the narratives.

Section 9.5 (Analysis of management rates). This section looks at how well the threats, variations and UOSs were managed, by looking at the percentages of each that were managed mismanaged or unmanaged. An analysis and examples are provided of the types of threats and variations that are most likely to be managed. Finally, a comparison is offered between the management rates recorded using the NOM tool on the airport ramp and management rates recorded from other LOSA programs, as well as the LOSA collaborative database.

Section 9.6 (Summary and Discussion of the Data). This section offers a preliminary discussion of the data collected from the observation forms and narratives, including the threats, variations and UOSs overall. The relationships between threats, variations and undesired operational states is discussed in Chapter 10. An evaluation of the NOM method is provided in Chapter 11.

9.1 Data Collected from Observation Forms

A total of 1302 observations were included in the analysis. For each of these, the observation forms (described in Chapter 8) collected demographic data about the type of port (regional, domestic or international), the specific location (airport, bay number), the type of bay (stand-off, aerobridge etc.) and the aircraft type. Observation forms also recorded the team size, position of the observer, and whether or not the aircraft departed on time. This section provides a high-level overview of the data collected from observation forms.

9.1.1 Ports visited

Table 6 below shows the distribution of observations across domestic and international ports. The international airports visited included airports in Argentina, Australia, Germany, India, Indonesia, South Africa, Thailand, the South Pacific, the United Kingdom and the United States of America.

Most of the observations (82.4%) were conducted in large domestic ports; 11.3% were conducted at international ports and 6.1% in domestic regional ports.

Table 6: Number and percentage of observations conducted at International and Australian Domestic Ports visited.

Port Type	Ports visited	No. of observations	% of total observations
Domestic ports observed:	**16**	**1154**	**88.6%**
Regional Ports	8	80	6.1%
Large Domestic Ports	8	1074	82.4%
International ports observed	**12**	**148**	**11.3%**
Total	**28**	**1302**	

9.1.2 Aircraft types

There were 10 aircraft types observed, including Airbus, Boeing and Bombardier Aerospace aircraft as shown in Table 7.

Table 7: Number of observations conducted for each aircraft type.

Aircraft Type	B737	B767	B744	A330	B17	A380	D8	A320	B747	B777
No. of obs.	769	236	144	56	37	26	16	8	8	2
% of total obs.	59.06	18.12	11.05	4.30	2.84	1.99	1.22	0.61	0.61	0.15

9.1.3 Aircraft observation area

Some aircraft, such as the DH8, only have one loading area, the aft hold at the back of the aircraft. Others, such as the B737, have two loading areas, one at the back and one at the front of the aircraft. Observers recorded the main location of each observation as taking place mainly at the back of the aircraft (aft hold) or mainly at the front of the aircraft (forward hold). The location was recorded as 'general' if the location of the observer was not restricted to any particular part of the aircraft. As shown in Table 8, just over three quarters of the observations fell into this category. Of the remainder, around two thirds were in the aft hold and one third in the forward hold. Table 8 also shows the number of observations conducted in each area of the aircraft; most observations were conducted in the general area in and around the aircraft.

Table 8: Aircraft type and the area of the aircraft observed.

Aircraft type	Aft Hold	Forward Hold	General	Total	% of all observations
A320	2	0	6	8	0.6%
A330	4	2	50	56	4.3%
A380	0	0	26	26	1.9%
B717	13	1	23	37	2.8%
B737	150	116	503	769	59.1%
B744	2	5	137	144	11.1%
B747	2	1	5	8	0.6%
B767	35	18	183	236	18.1%
B777	0	0	2	2	0.1%
DH8	16	0	0	16	1.2%
Total	224 (18.6%)	143 (11.91%)	935 (77.91%)	1302 (100%)	100%

9.1.4 Activity observed

Turnarounds generally involve unloading the arriving aircraft and loading it again for departure. However, each turnaround was split into loading or unloading observations to allow for initiating and terminating flights. Therefore, where the observer observed the whole turnaround, this would count as two observations. As shown in Table 9, the overall numbers of unloading and loading observations were roughly equal.

Table 9: The number of observations conducted for loading and unloading activity.

Loading activity observed	No. of observations	% of total observations
Loading	660	50.7%
Unloading	642	49.3%

Total	1302	100%

9.1.5 Time of day

The observation start time was recorded from the aircraft's arrival at the bay and the finish time was the aircraft's departure. Most observations were conducted within a 12-hour period, between 06:00 and 18:00, with the least occurring in the 00:00–06:00 period. This reflects the period of time that most of the ports were operational, as some ports operate a curfew between 23:00 and 06:00.

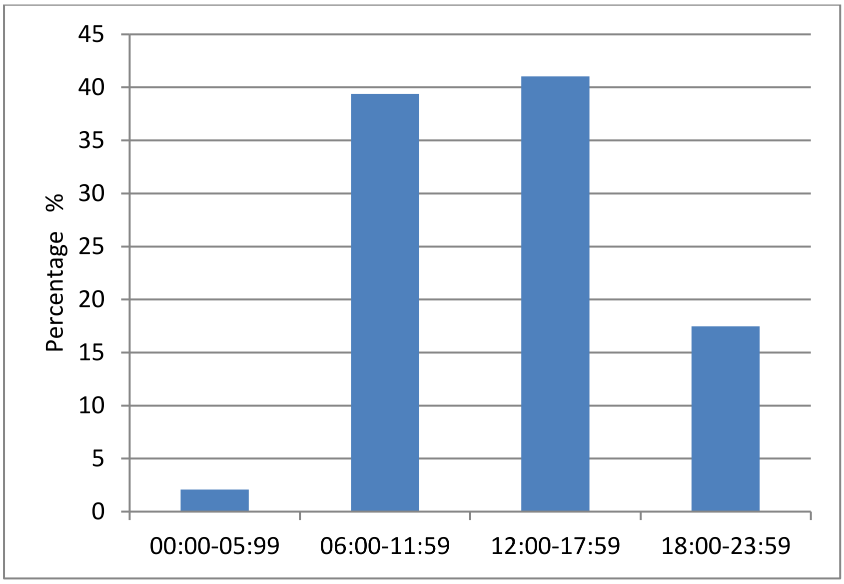


Figure 17: Percentage of observations conducted at within each time period.

9.1.6 Ground crew size

The most commonly observed crew size had five-member teams (42% of all observations). Teams of six members were also commonly observed, whereas only 14% of observations recorded crews of four or less. Very few observations recorded crews of seven or more. Figure 8 represents the percentages of the different team sizes observed.

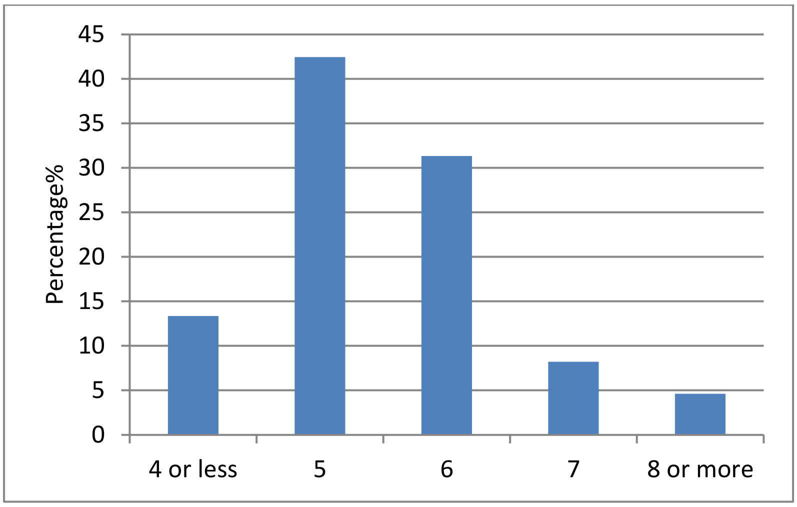


Figure 18: Percentages of different team sizes observed.

Crew size varies between ports and in relation to the aircraft size and type. The maximum team size recorded was 10, the minimum was two. Teams of five and six members were the most commonly observed (see Figure 20).

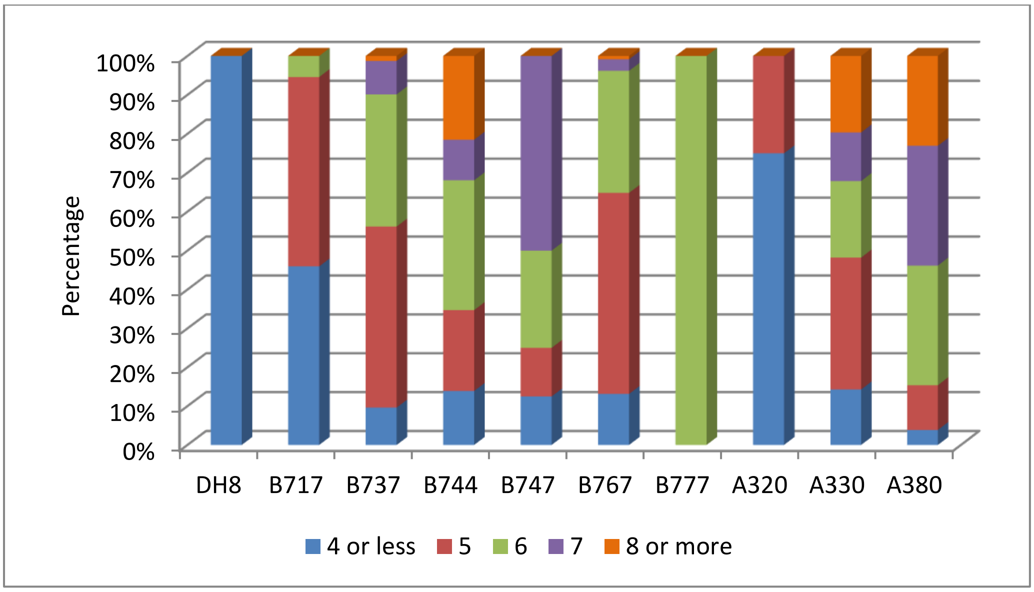


Figure 19: Percentage of observations for each aircraft type by crew team size.

Larger crews are typically assigned to larger aircraft in busier ports or with larger cargo payloads. Figure 19 shows smaller aircraft, like DH8s, A320s and B717s, tended to have crews of 4–5 or fewer, whereas larger aircraft such as A380 and B744 were assigned crews of more than five. Factors other than aircraft size, however, also seemed to impact on the size of the crew assigned, since smaller aircraft like A330s also had the largest crews, and very large aircraft like B767s and B777s often did not.

9.1.7 Arrival/departure time and delays

Each observation recorded the aircraft's arrival and departure time, as well as whether this was as scheduled, early or late (see Figure 2020).

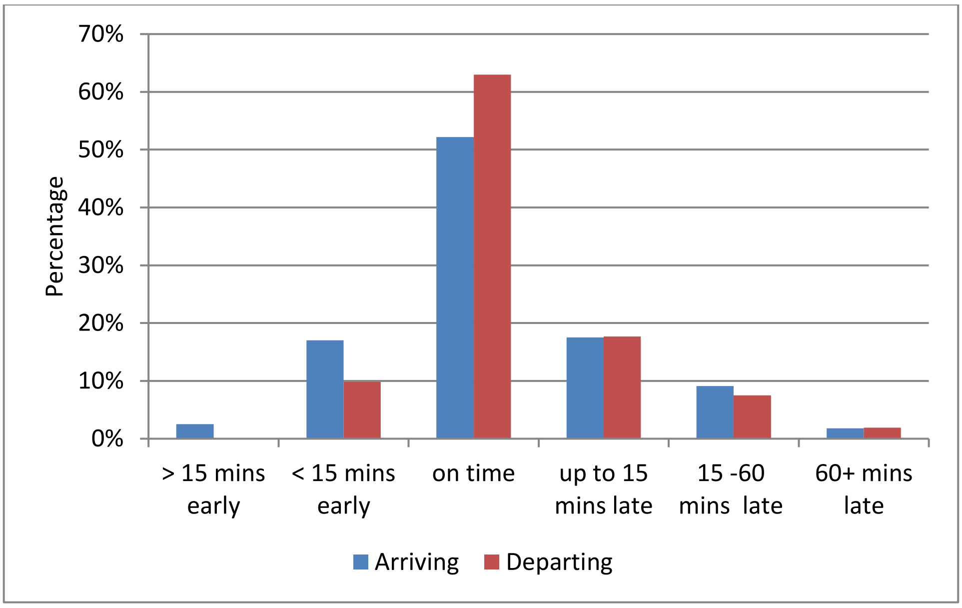


Figure 20: Percentage of observations in which the aircraft arrived on time, early or late.

At least half the aircraft observed were on time, with the remainder more likely to be late than early. The arriving (in blue) and departing aircraft (in red) show a similar pattern.

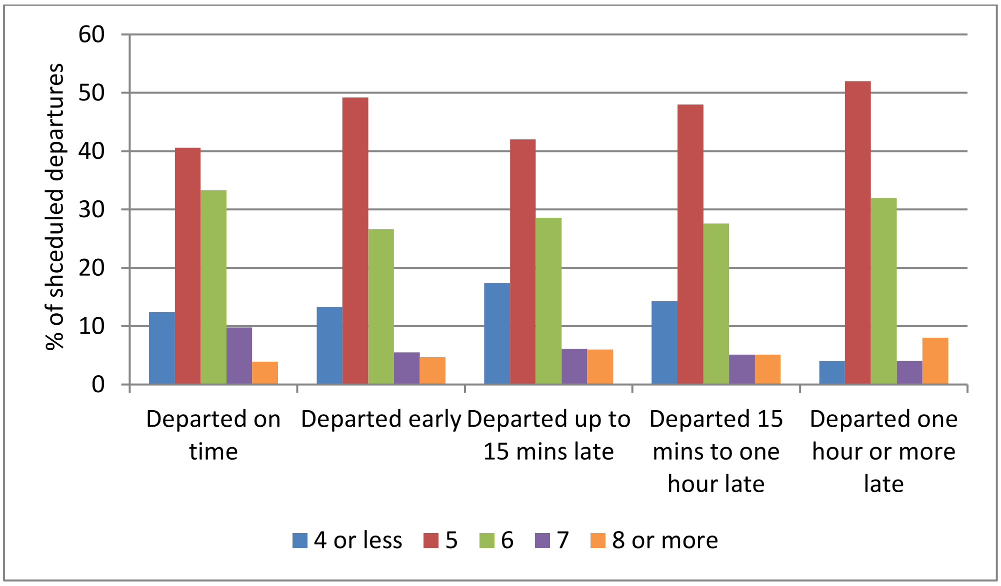


Figure 21: Percentages of early/on time/ late services for different crew sizes.

Figure 2121 shows the percentages of services that departed on time, early or late, based on the number of crew in the ground servicing team. On time, early and late departures showed a similar pattern for each crew size.

9.1.8 Normal, abnormal and emergency modes

The NOM Model in described in Chapter 8 describes the three possible operational modes.

Table 10: Observations conducted in each operational mode.

Operational mode	Number of Observations	% of Observations
Normal Operations	1137	87%
Abnormal Operations (with a UOS)	164	12.6%
Emergency Operations (with a recordable incident)	1	0.7%
Total number of Obs	1302	100%

The majority of the observations were of normal operations (87%) as shown in Table 10, with 12.6% observing abnormal conditions. There was only one observation that recorded a UOS that resulted in a recordable incident.

9.2 Analysis of Threats

9.2.1 Number of threats recorded in each observation

There were over 2437 threats coded from the total of 1302 observations conducted. Most observations (87.4%) recorded at least one threat. The highest number of threats recorded in any one observation was 10. Figure 222 below shows the percentage of observation periods with one or more threats. Half the observations had a least two threats and a further 27% had 3–4. Very few observations had 8 or more threats. There were 160 observations (12.6%) where no threat was recorded.

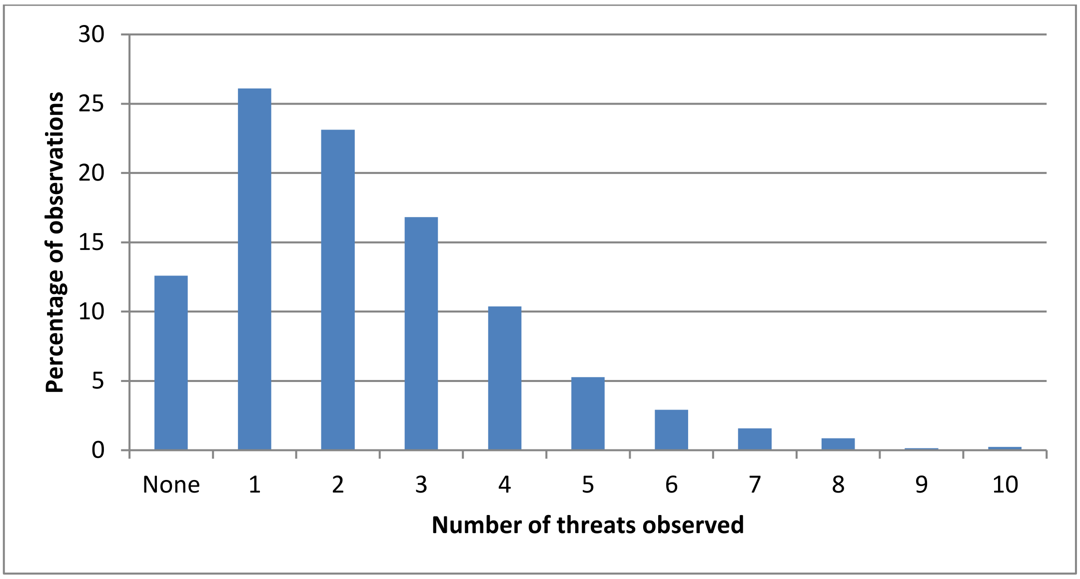


Figure 22: Percentage of observation periods with one or more threats observed.

9.2.2 Overview of major threat categories

Threats are grouped into 13 major categories. Table 11 shows the distribution of threats occurring in each major category and the percentage of observations that saw a threat of this type. The mean was 1.87 threats per observation. The highest number of threats occurred in the 'Airside Driving' category, followed by 'Equipment' and 'Foreign Object Debris' threats (shown in bold below).

Table 11. The distribution of threats in each major threat category.

Threat major category	No. of sub-categories within major category	Total No. of threats	% total threats	No. of obs. with this threat	% obs. with this threat
1. Airport facilities	7	169	6.4%	148	11.4%
2. Airside driving	**15**	**922**	**31.0%**	**680**	**52.2%**
3. Arriving irregular load	7	64	2.2%	53	4.1%
4. Bay management	4	27	0.9%	26	2.0%
5. Communication	7	124	4.2%	120	9.2%
6. Documentation	4	49	1.7%	48	3.7%
7. Equipment	**11**	**534**	**18.3%**	**447**	**34.3%**
8. Foreign Object Debris	**2**	**395**	**13.5%**	**367**	**28.2%**
9. Freight/ Baggage	22	385	13.3%	309	23.7%
10. Load Control	7	84	2.7%	78	6.0%
11. Manpower	8	36	1.2%	30	2.3%
12. Operational pressures	10	116	4.0%	113	8.7%
13. Weather	6	18	0.6%	18	1.4%
14. Obs. With no threat	0	0	0	160	12.6%
Total	112	2923	100%	*	**

*Total Observations = 1302.
** does not sum to 100 % as each observation can have more than one threat type.

9.2.3 Subcategory threats

The 13 major threat categories shown in Table 11 above have been broken down into subcategories in the figures below to show, in more detail, the individual threats contributing to each major category.

1. Airside Driving subcategory threats

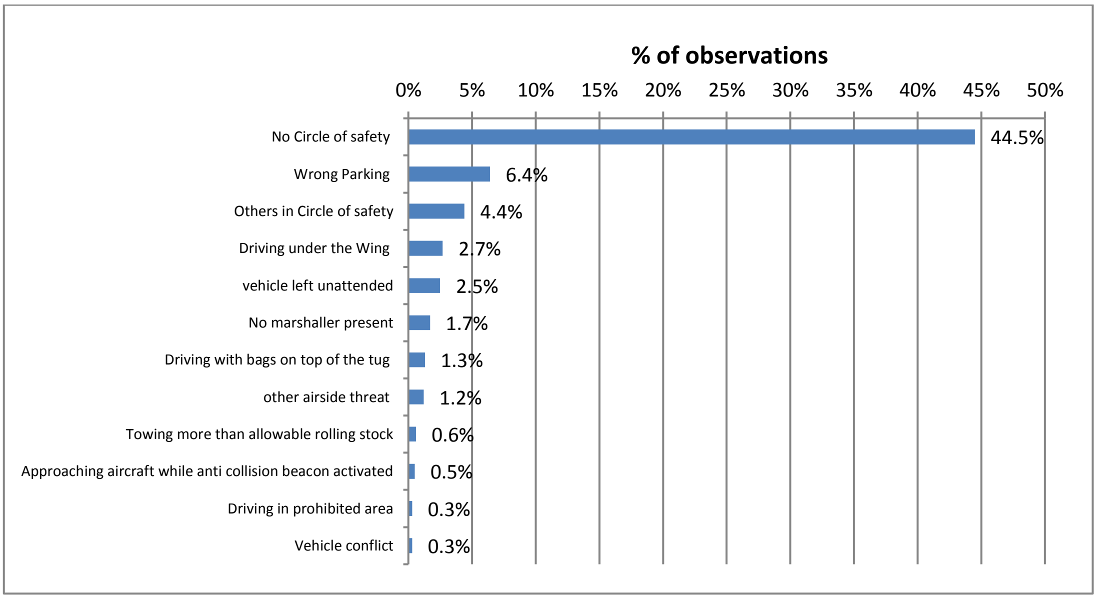


Figure 23: Percentage of each airside driving threat subcategory occurring across all observations.

The most frequently occurring major category was airside driving threats (22). There were 680 observations with at least one airside driving threat. **Error! Reference source not found.**3 shows the subcategories contributing to this this group. Nearby personnel (not part of the ramp crew) failing to perform the circle of safety was by far the most frequently occurring airside driving threat, affecting over 44% of all observations recorded. Parking in the wrong place (e.g. outside of the permitted area) was the second most frequent threat, affecting 6.4% of observations.

2. Equipment threats

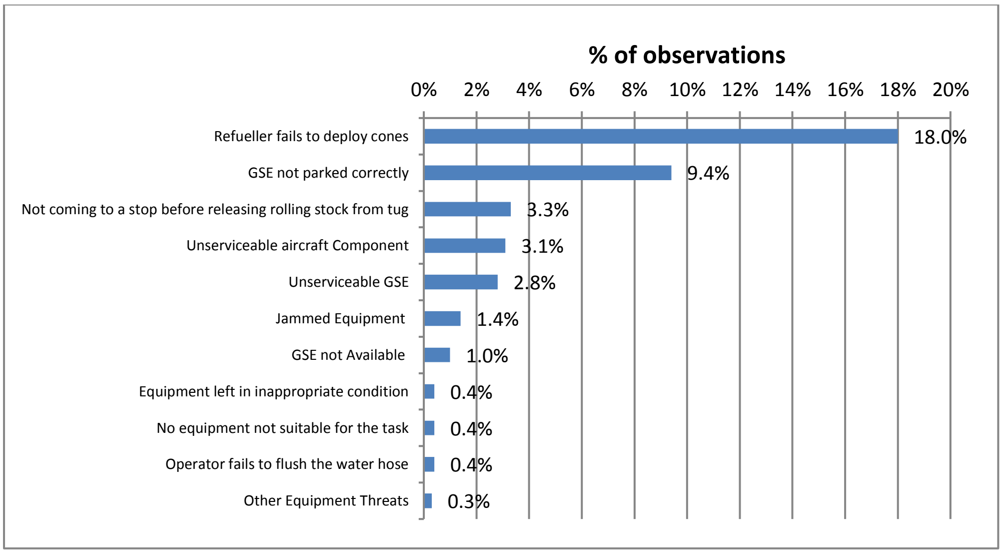


Figure 24: Percentages of each Equipment threat subcategory occurring across all observations.

Equipment threats were the second highest major category (44). Figure 244 shows the subcategories within Equipment threats. A third of observations (477) had at least one equipment threat. About half of these (18%) were the refueller failing to deploy cones around fuel hoses during the refuelling process. This was followed by the incorrect parking of ground servicing equipment (GSE), which occurred in 9.4% of observations.

3. Foreign Object Debris (FOD) threats

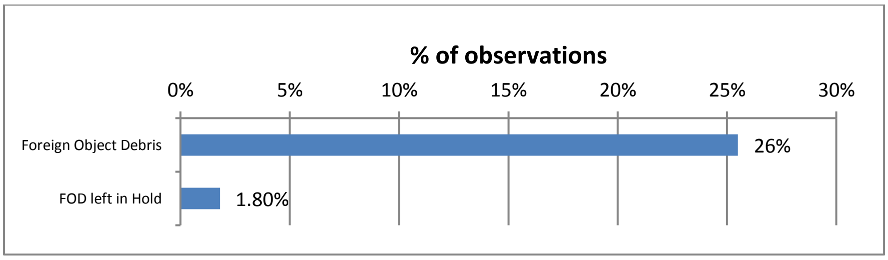


Figure 25: Percentages of each FOD threat subcategory occurring across all observations.

There were 395 observations with at least one threat of foreign object debris (FOD), with 28.2% of observations recording this threat. The most common occurrence was FOD left on the ramp (26%); however, there were a few occasions where FOD was also found inside the aircraft hold (1.8%).

4. Freight and Baggage threats

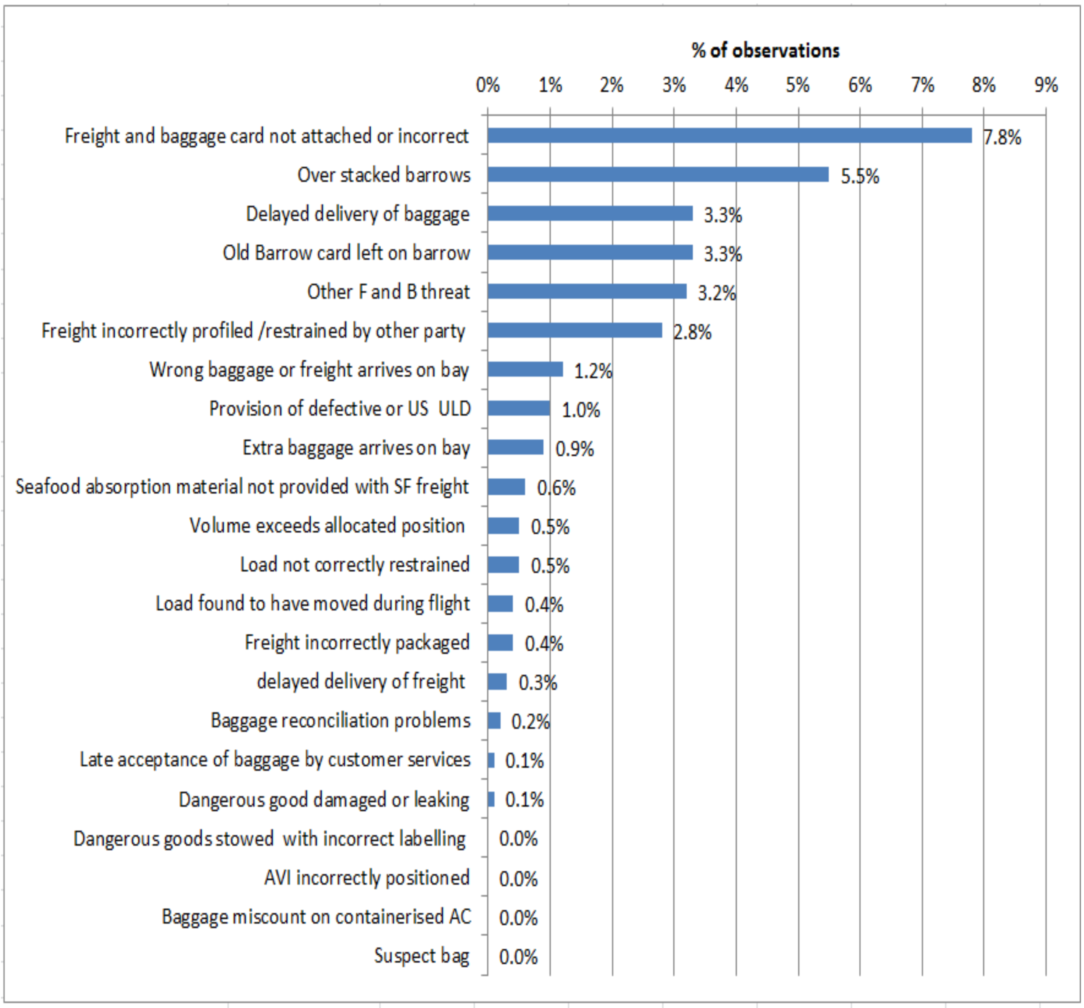


Figure 26: Percentages of each Freight and Baggage threat subcategory occurring across all observations.

There were 395 Freight and Baggage threats. The most common was freight and baggage presented without the correct label cards attached (7.8%); barrows were also over-stacked (5%), had old barrow cards (3.3%) or were delivered late (3.3 %).

5. Airport Facility threats

There were 169 Airport Facility threats related to the airport facility itself, such as the bays, roads, parking areas or lighting. There were 148 observations that recorded at least one airport facility threat. As shown in Figure 27, the most commonly occurring was Ground Service Equipment (GSE) congestion in and around the bay, occurring in 5.9% of observations.

The second most common threat was the lack of parking or staging areas for GSE, which occurred in 1.4% of observations.

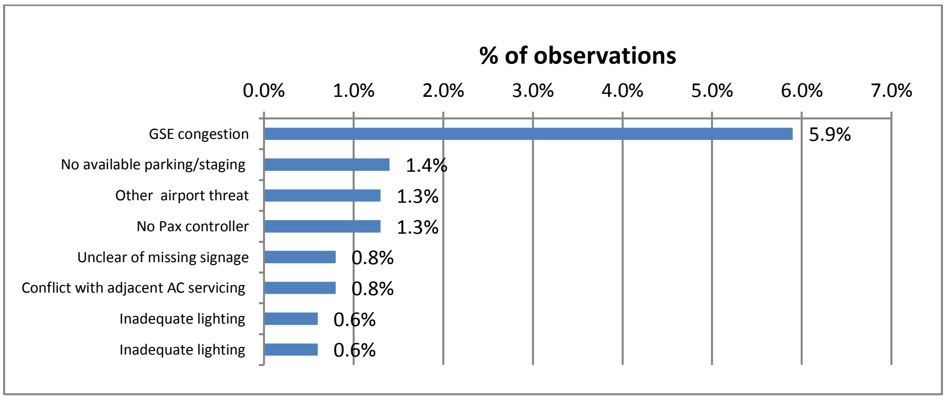


Figure 27: Percentages of each Airport Facility threat subcategory, occurring across all observations.

6. Communication threats

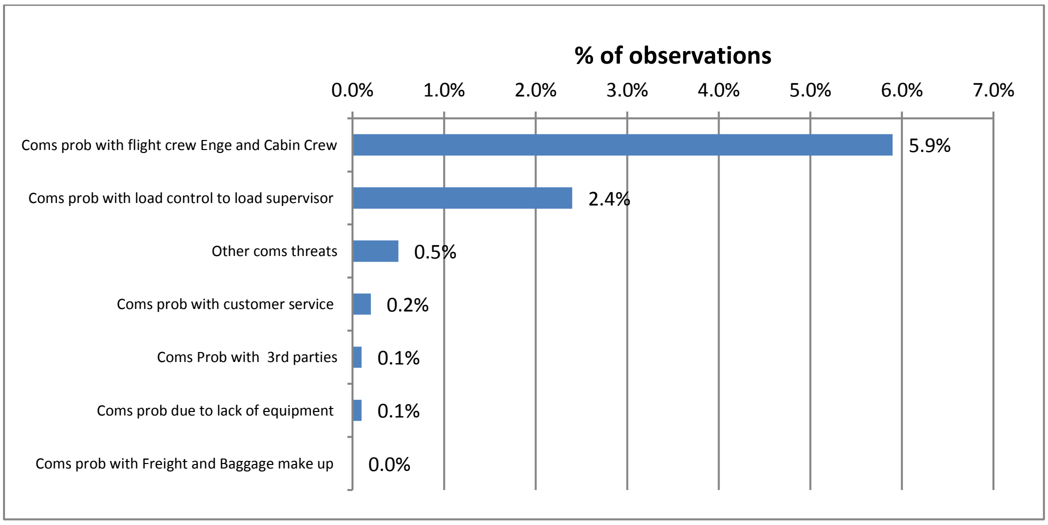


Figure 28: Percentages of each Communication threat subcategory occurring across all observations.

A total of 124 communication threats were recorded across 120 observations. Figure 288 shows 5.9 % of observations were affected by communication problems occurring between the flight crew engineer and the cabin crew. Communication problems also occurred between Load Control and the Loading Supervisor, affecting 2.4% of observations.

7. Operational pressures

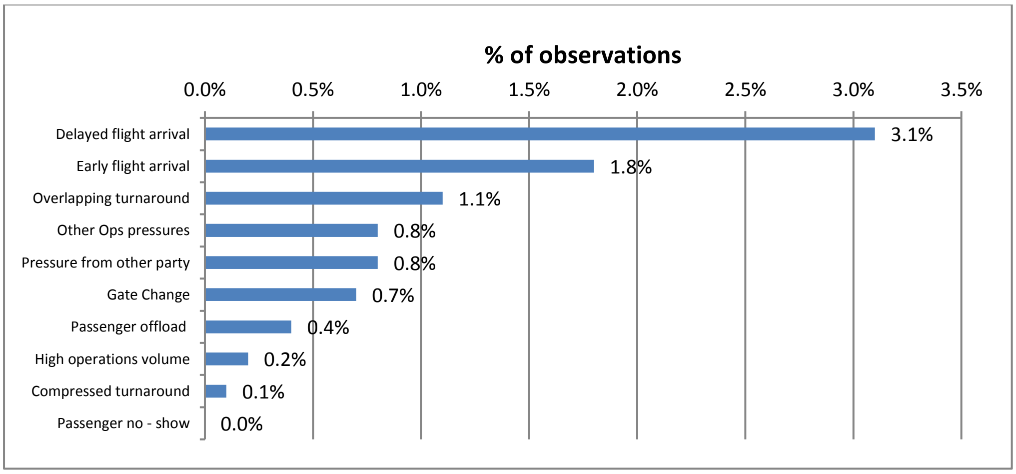


Figure 29: Percentages of each Operational pressure threat subcategory occurring across all observations.

There were 116 operational pressure threats. There were 113 observations with at least one of these threats. Figure 2929 shows delayed flight arrival affected 3.1 % of observations, resulting in reduced time to complete the turnaround tasks. Early flight arrival affected 1.8% of observations, where crew may have had less time to prepare for the aircraft's arrival. Overlapping turnarounds was also an issue in 1.1% of observations, where the time available for one turnaround overlapped with the time scheduled for the next turnaround.

8. Load Control threats

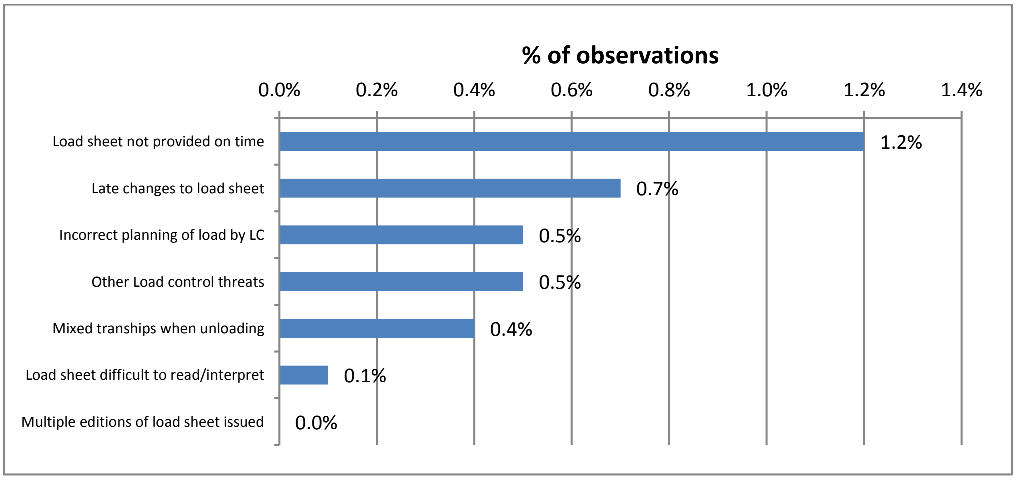


Figure 30: Percentage of each Load Control threat subcategory occurring across all observations.

There were 84 Load Control threats recorded and 78 observations with at least one Load Control threat. Figure 30 shows the load sheet or Load Instruction Report (LIR) not being provided on time in 1.2% of observations. Late changes to the load sheet were observed in 0.7% of all observations.

9. Arriving Irregular Load threats

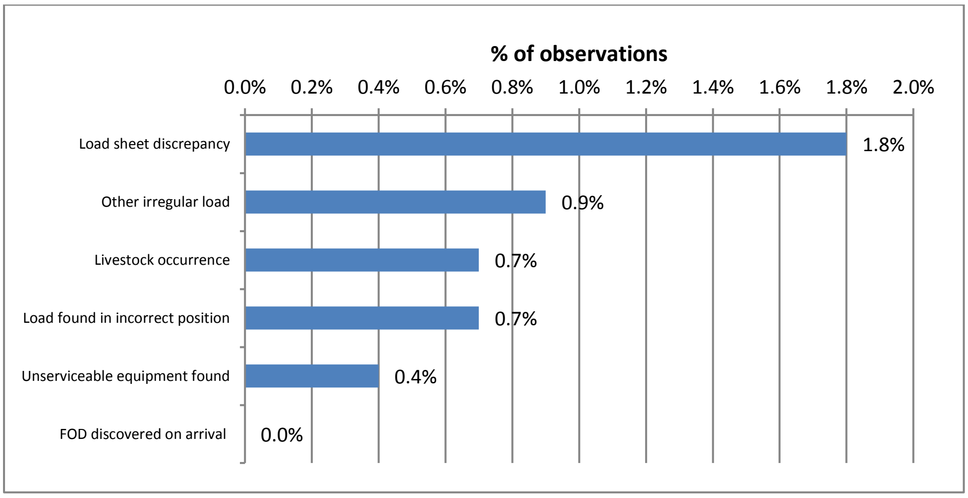


Figure 31: Percentage of each of Arriving Irregular Load threat subcategory occurring across all observations.

There were 64 Arriving Irregular Load threats recorded across 53 observations. Figure 3131 shows a large portion (1.8%) were load sheet discrepancies, where the load or cargo did not match the load sheet (or Load Instruction Report). 'Other irregular loads' affected 0.9% of all observations.

10. Documentation threats

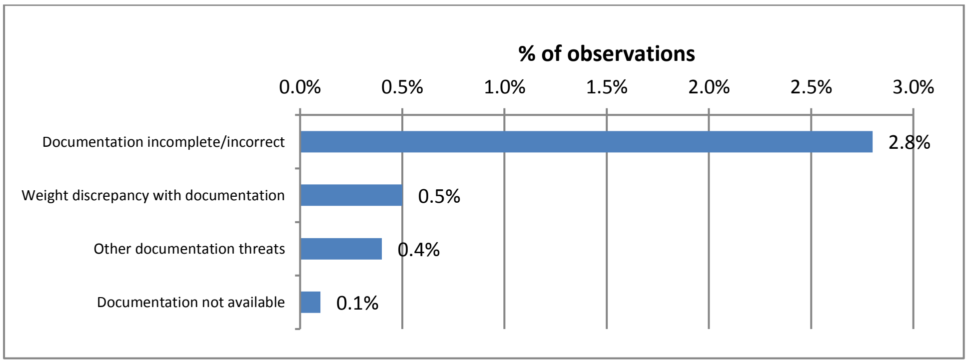


Figure 32: Percentage of each of Documentation threat subcategory occurring across all observations.

There were 49 Documentation threats observed over 48 observations. Figure 322 shows the most frequently occurring documentation threat was incomplete or incorrect documentation given to the ramp crew, affecting 2.8% of observations, followed by weight discrepancy threats within the documentation, which occurred in 0.5% of observations.

11. Manpower threats

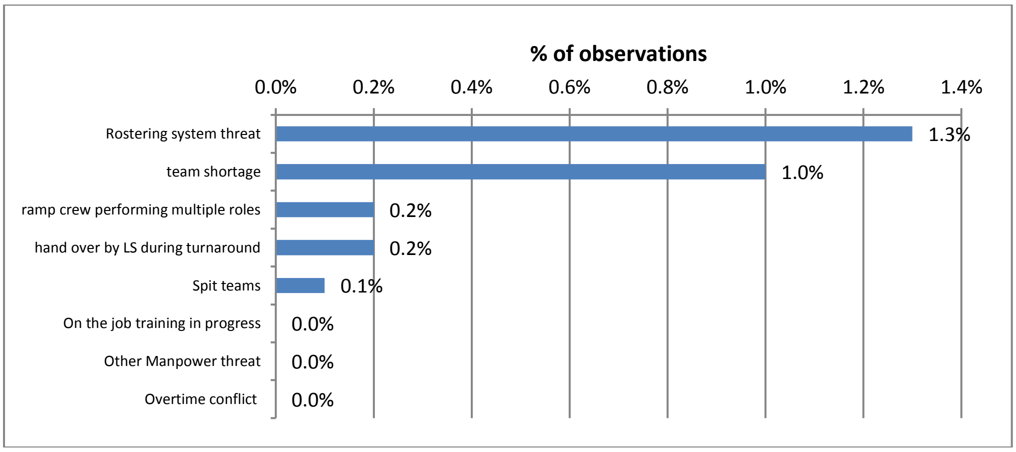


Figure 33: Percentage of each of Manpower threat subcategory occurring across all observations.

Manpower threats refer to the number of crew available to complete the turnaround. There were only 36 manpower threats recorded across 30 observations. Figure 333 shows the largest occurring subcategory was rostering threat, occurring in 1.3% of observations; 1% of observations were also affected by team shortages.

12. Bay Management threats

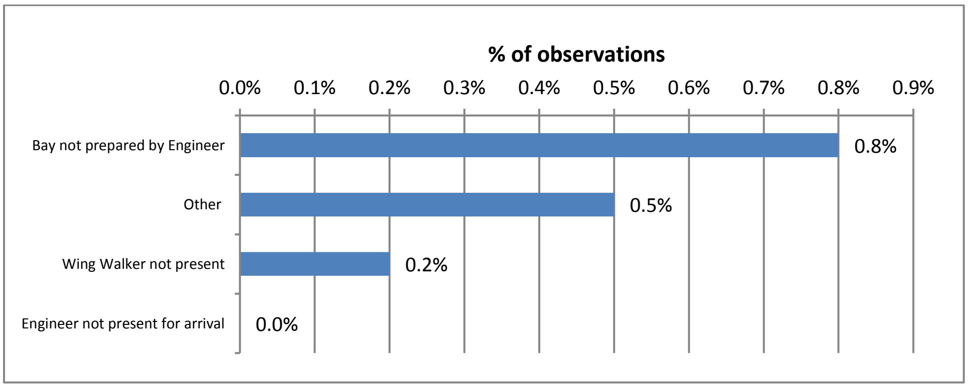


Figure 34: Percentage of each of Bay Management threat subcategory occurring across all observations.

There were 27 Bay Management threats recorded within 26 observations. Figure 344 shows the most common was the bay not being properly prepared by the Engineer, affecting 0.8% of observations, followed by 'other' Bay Management threats, affecting 0.5%.

13. Weather threats

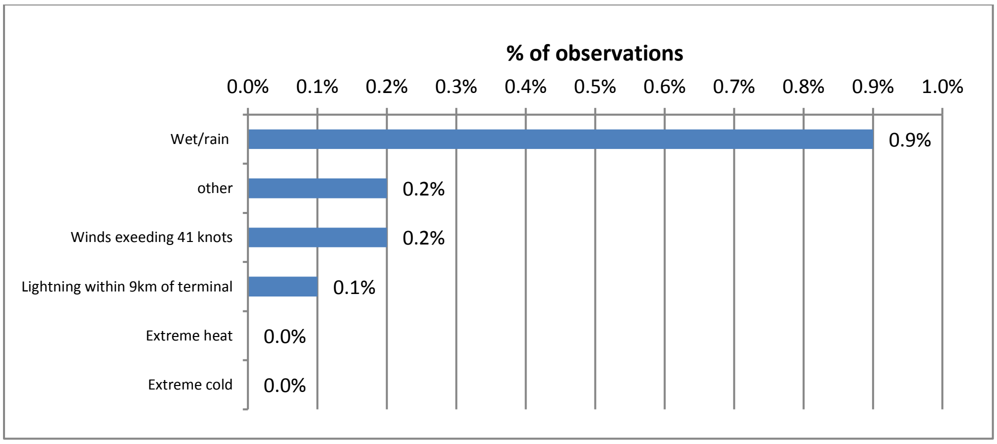


Figure 35: Percentage of each of Weather threat subcategory occurring across all observations.

Weather threats overall were much lower than expected, with only 18 observations recording any weather threat. However, the data does not reflect when operations were suspended due to adverse weather conditions, such as lightning or hail. Figure 355 shows 0.9% of these threats were attributed to rainy conditions. 'Other' weather threats affected 0.2% of observations, suggesting additional subcategories may be needed to describe all possible weather conditions.

9.2.4 Most frequent subcategory threats

Some threat types were observed much more commonly than others. The five most frequently occurring threats are shown in Table 12. Examples of each threat type are provided in the descriptions below.

Table 12: The most frequently occurring threats.

Top 5 sub-category threat types	Major threat category	Total no. obs with threat	% of obs affected
1. No circle of safety performed (Non-ramp crew)	Airside Driving Threat	579	44.5%
2. Foreign Object Debris on ramp	Foreign Object Debris Threat	345	26%
3. Refueller fails to deploy witches hats	Equipment Threat	234	17%
4. Ground servicing equipment not parked correctly	Equipment Threat	122	9.4%

5. Correct cards/label not attached to the load	Freight and Baggage Threat	101	7.8%

1. **No circle of safety performed.** Almost half of all observations included the Airside Driving Threat of 'no circle of safety performed'. Non-compliance with this procedure was the most frequently occurring threat, occurring in 44.5% of observations. Example from narrative:

   23:03 a catering truck arrives and stops to let the marshaller out, the truck is then marshalled into the forward cabin door without performing the Circle of Safety.

2. **Foreign Object Debris on ramp.** The second most commonly observed threat of all was leaving Foreign Object Debris (FOD) around the aircraft, occurring in 26% of observations. Example from narrative:

   Once the JCPL was in position the Auditor inspected the bulk hold. The bulk hold contained numerous brown lines, old LIRs and rubbish from bags. This was not removed from the bulk hold prior to departure of the aircraft.

3. **Refueller fails to deploy 'witches hats'.** The third most common threat was from the major category of Equipment Threats involved in refuelling the aircraft – failing to protect the fuel hoses with orange traffic cones. This occurred in 17% of observations. Example from narrative:

   A refuelling truck was connected to the aircraft on the port side and there was a safety cone positioned at the coupling on the ground. There were no safety cones positioned around the fuel hose on the ground.

4. **Ground servicing equipment not parked correctly.** The fourth highest threat was the Equipment Threat of failure to park ground servicing equipment in the designated areas, occurring in 9.4% of observations. Example from narrative:

   The aircraft began to taxi onto the bay and the belt loader was still positioned outside the equipment clearance line.

5. **Correct cards/label not attached to the load.** The fifth most frequently occurring threat was from the Freight and Baggage category. This threat involved incorrect or lack of labelling of freight, which can result in the wrong freight being loaded or freight being loaded into the wrong position. This was observed in 7.8% of observations. Example from narrative:

   The driver then moved forward and the last baggage container, containing crew bags, was then loaded onto the rear platform. It was noted that there was no ULD card in this container.

9.2.5 Where threats are occurring

Table 13: Mean threats per observation for each port type.

Port Types	N	Mean	Standard Deviation
International	290	3.46	1.93
Domestic	774	2.38	1.43
Regional domestic	48	1.6	0.93

The mean number of threats recorded during any one observation was significantly higher at international ports (SD 1.939), as shown in Table 13.

9.2.6 Threats and crew size

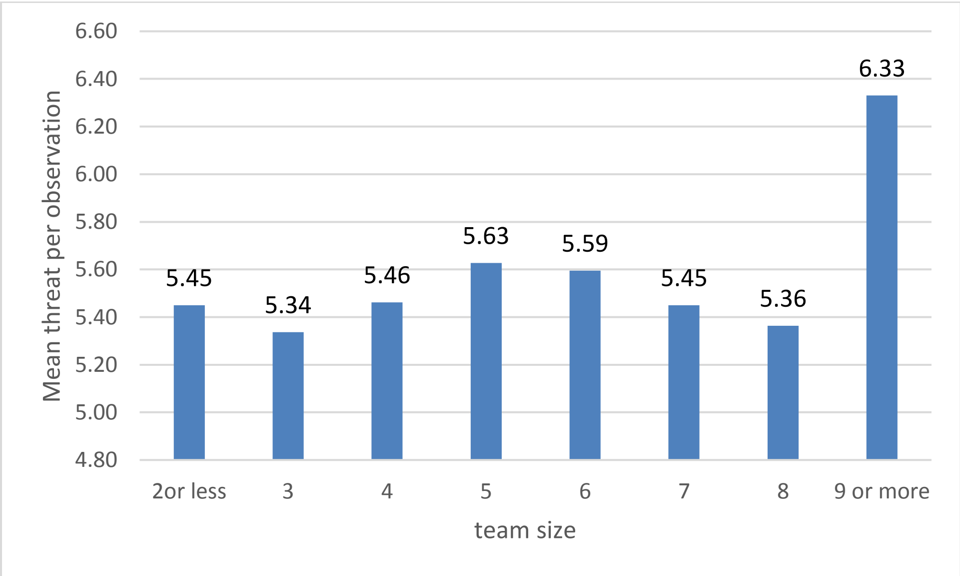


Figure 36 Mean threats recorded for each crew size.

Figure 366 shows that most team sizes had minimal impact on the mean threats per observation, except where there were 9 crew or more. Having a team size of 9 or more led to a higher than average mean threat, of 6.3 per observation as shown in Figure 366.

9.2.7 Threat origins – other airport teams

Observers were asked to indicate where threats originated, if known. In some cases, the origin of the threat was not observable (427 threats for example, were coded as unknown). Table 14 indicates many threats (19% of all observations) originated from other ramp crew who were not part of the turnaround being observed but were servicing nearby aircraft. A large number of threats (17.5 %) also came from catering teams carrying out tasks in the same vicinity as the ramp crew. Threats from the Engineering teams featured in 9.2% of observations. The majority of the catering threats are driving issues, while threats originating from the Freight include poorly prepared freight and baggage arriving on the bay. Examples of the types of threats observed from each area are shown in Table 14.

Table 14: Examples of the types threats that can originate from other airport teams

Other Airport Team	Obs. with threat inv. this dept.	% obs. affected	Examples of typical threats originating from each airport team extracted from narratives
Other Ramp teams	584	19.0	*driver takes short cut through servicing area.*
Catering	533	17.5	*There is congestion at the forward hold with the Catering truck and the FMC both in position. The tug driver finds it hard to line up the low profile to the rear of the FMC to unload a pallet due to the position of the catering truck.*
Freight	292	9.6	*A freight delivery driver drops a barrow in the storage area, it contains an AVI and an AOG part; there is no barrow card in the pouch on the front of the barrow.*
Engineering	281	9.2	*Engineer had moved his buggy to the front of the aircraft which had blocked access to the forward cargo door. When the loading supervisor arrived with the belt loader the engineer was not in the area and could not be found to move his buggy, the forward hold operator had tried to move the buggy but he did not know how to operate it. The engineer had returned to his buggy and moved the buggy out of the service area but it had already caused a delay for the ramp staff to start loading the aircraft.*
Refuellers	245	8.0	*A fuel tanker approached the aircraft and hooked up to the fuelling point beneath the wing, the operator failed to position any witches hats adjacent to the fuel hose as required to warn drivers of its position. I look around the truck and fail to find any witches hats on the vehicle or a place to store them.*
Baggage Room	160	5.2	*Baggage room driver drops another barrow of bags at the bay; he does not come to a complete stop before disconnecting it from the barrow. The barrow rolls for approx. 1 meter before stopping.*

Other Airport Team	Obs. with threat inv. this dept.	% obs. affected	***Examples of typical threats originating from each airport team, extracted from narratives***
Load Control	135	4.4	*The dry ice was identified, but not as a 'Dangerous Goods' on the Load Instruction Report. Load Control should not have planned the dry ice to be loaded in the FWD hold with AVIs.*
Airport Owner	116	3.8	*The lighting on this bay is very dim. I moved back to the E&M shed to see my paper work. Lighting is inadequate and gets progressively worse the further away the bays are from the terminal.*
Other Airports teams	97	3.2	*Vehicle left unattended with the engine running.*
Flight Operations	84	2.8	*I arrived on [the bay]. There was already aircraft parked on this bay waiting for push back clearance. The aircraft [waiting] for this bay was parked on a stand-off bay and was ready to be towed and parked on [to the bay] once the bay was clear.*
GSE	77	2.5	*The belt loader that was used at the front has a faulty braking system and the staff used a chock to keep the belt in position.*

9.2.8 Other possible analysis of threats

The data collected regarding threats allows for analysis in a number of different ways, not all of which can be included here. For example, it is possible to analyse the data around any particular threat in terms of geographical location, airport bay, aircraft type or number of crew. It is also possible to use the qualitative information from narratives, observer issue fields or port reports to provide additional context to threats occurring. Examples of how this data might be further analysed are shown below.

Example analysis of threat origin. It is possible to analyse any particular threat by location or based on the team from which the threat originates. Figure 377 (over), for example, shows the breakdown of threat origins by other airport team for the threat of Ground Service Equipment not parked correctly.

The figure shows that the Engineering crew are responsible for the largest number of parking threats, where personnel from the Engineering teams have parked equipment or vehicles in areas not designated for parking or staging. This is interesting because there are typically fewer engineering vehicles on bay than vehicles from other departments, suggesting a small number of personnel may be contributing to a large proportion of threats.

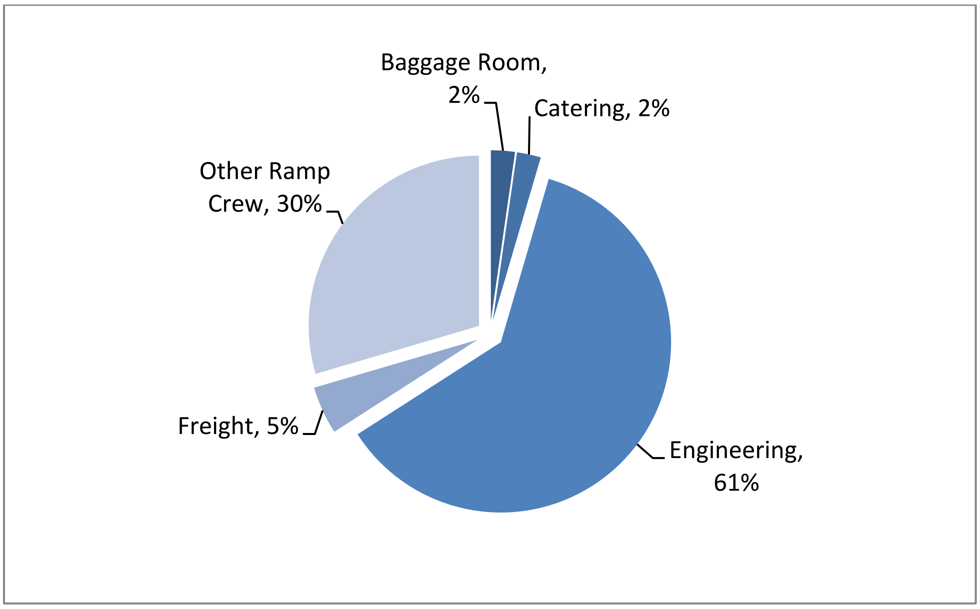


Figure 37 Breakdown of GSE parking threats by area

Example analysis of observer issues. As described in Chapter 8, observers were able to add additional information that they think may be relevant in the 'observer issues' field in the data base. Observer issues provided context about procedures that are ambiguous, difficult to comply with, or had been implemented differently in some locations. Examples of observer issues relevant to the coding of threats are provided below:

Marshalling procedures**:** '*there is confusion about when a marshaller is required due to contradictions in the ramp manual. Depending on which rule is referred to for the use of marshalling for stairs, it could be interpreted as mandatory or discretionary'.*

Refuelling procedures**:** '*The procedure for fuelling by hydrant dispenser states that: To make drivers more aware of the location of fuelling hoses and coupling, cones and flags are used to indicate their position. However the procedure is not mandatory for refuellers. The high level of refuellers not using cones noted during the wash phase lead to the concern being raised about mismatch between Ramp and Refuelling procedures. There may be an assumption by ramp crew that cones will always be deployed but this is not the case'.*

Procedures for barrow cards**:** '*Old baggage cards left on containers have the potential to cause confusion and may contribute to loading errors if incorrect labels are referred to. There is no* *documented procedure outlining who has the responsibility to remove old barrow cards. It is different in each port, for instance Loading Supervisors remove old barrow cards in some ports whereas in others barrow cards should be removed by the last personnel to handle the barrows. Failure to remove barrow cards has the potential to cause confusion with the loading process, however this could not be coded as an error as it was not a written procedure for Ramp crew'.*

The excepts above are provided as examples only to demonstrate that further insights can be gained from the analysis of observer issue related to the occurrence of threats.

Example analysis of port reports. Observers were asked to complete port reports regarding any local conditions in each port that might be relevant to the interpretation of data in that location. Example of port reports for different locations provided below:

Location specific congestion threats: *'The domestic bays 1 and 2 are both bays for non-containerised aircraft with fixed aerobridges that are located extremely close to the terminal. Bay 1 has a fixed concrete plinth that is located near the entry path for the catering truck to the fwd. galley; this plinth does not allow the catering truck to position correctly which restricts ramp manoeuvring area. Bay 2 has high congestion of GSE parked within and outside equipment limit lines due to the location of the ramp meal room. Bay 1 seems to be a thoroughfare for traffic that enters or leaves the baggage area'.*

'*Port M domestic bays 1 and 2 are both bays for non-containerised aircraft with fixed aerobridges that are located extremely close to the terminal'.*

'Port C Bay 1 has a fixed concrete plinth that is located near the entry path for the catering truck to the fwd. galley; this plinth does not allow the catering truck to position correctly which restricts ramp manoeuvring area'.

'Port A Bay 1 seems to be a thoroughfare for traffic that enters or leaves the baggage area and Bay 2 has high congestion of GSE parked within and outside equipment limit lines due to the location of the ramp meal room'.

These examples demonstrate how port reports also provide additional context for interpreting threats and identifying hypothesis for threat causation

1.3 Analysis of Variations

9.3.1 Number of variations recorded in each observation

There were over 3137 variations recorded across the 1302 observations conducted. **Error! Reference source not found.**8 shows most observations (93.6%) recorded at least one variation from standard operating procedures, with the maximum recorded in any one observation being 17. There were 81 observations (6.4%) with no variation recorded.

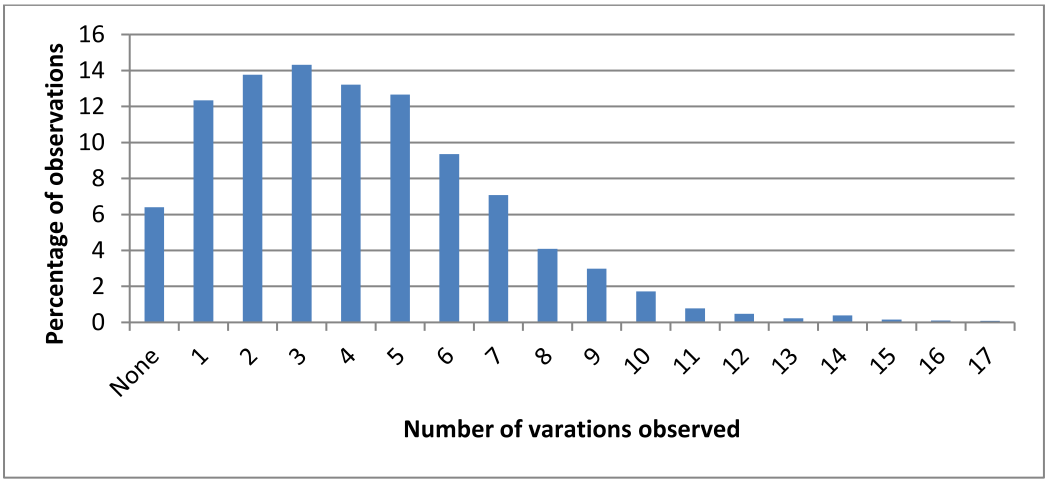


Figure 38 Percentage of observation periods with one or more variations observed.

9.3.2 Overview of major variation categories

Variations were grouped into 11 major categories;

Table 15 below shows the distribution of variations occurring in each major category and the percentage of observations that had a variation of this type.

Each observation could include more than one variation. The mean was mean = 2.43 variations occurring during each observation period.

Table 15: Variation category distribution.

Variation major category	No. of sub-categories within major category	Total no. of variations	% total variations	No. of obs. with this variation	% obs. with this variation
1. Bay management	7	109	3.23	102	7.8
2. Communication	4	17	0.54	17	1.3
3. Documentation	8	439	10.44	330	25.3
4. Door variations	5	119	3.67	116	8.9
5. Airside Driving	26	**2290**	**31.16**	**985**	**75.7**
6. Equipment usage	26	**1070**	**22.11**	**699**	**53.7**
7. Loading	16	359	7.05	223	17.1
8. Pushback	3	1	0.00	0	0
9. Read back	4	246	7.02	222	17.1
10. Restraining	6	289	8.54	270	20.7
11. Unloading	9	251	6.23	197	15.1
12. Obs. with no variation	0	0	0	81	6.4
Total	114	5190	100%	*	**

*Total Observations = 1302
** does not sum to 100 %as each observation can have more than one variation type

9.3.3 Subcategories of variations

Of the 11 major variation categories, the most prevalent were Airside Driving Variations, Equipment Usage Variations and Documentation Variations. Each major variation category has been broken down to identify contributing variation types below.

1. Airside Driving variations

There were 2290 Airside Driving Variations recorded across 985 observations (see Figure 39, over), making this the most common category of variations overall. The most frequently recorded Airside Driving Variation was the failure to perform the circle of safety, which was observed in 44% of observations. This was closely followed by leaving the vehicle engine running while unattended, which occurred in 43.3% of observations.

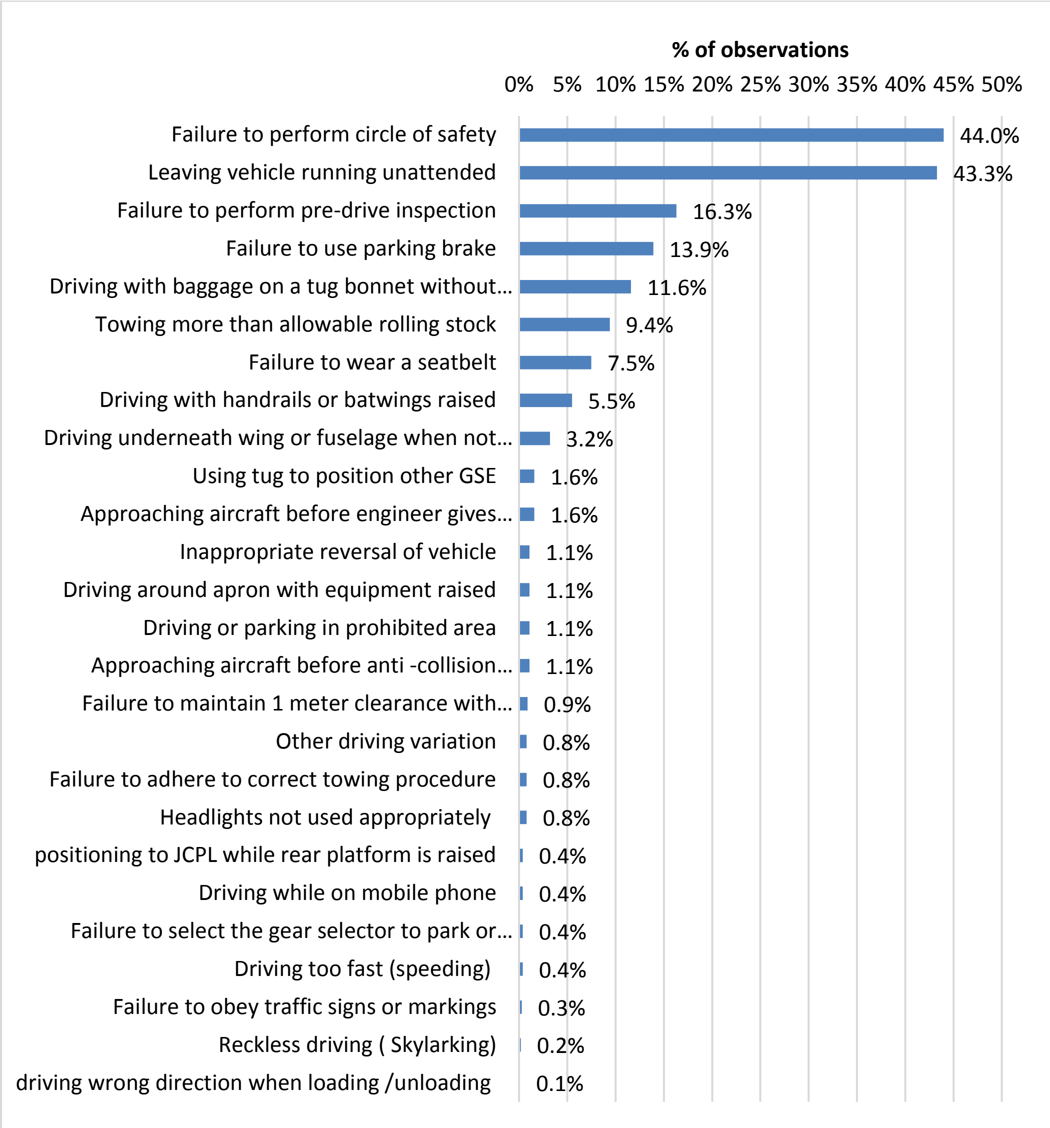


Figure 39 Percentage of each of Driving variation subcategory occurring across all observations.

2. Equipment variations

The next most common group of variations were Equipment usage variations. There were 1070 of these variations recorded across 699 observations. Figure 40 (over) shows that, of these, the most common was failing to put the control pendant in the hold while loading, affecting 28.8 % of variations, followed by the inappropriate deployment and retraction of handrails and batwings (retractable side rails) on belt loaders, which occurred in 7.5% of variations.

Figure 40 Percentage of each of Equipment variation subcategory occurring across all observations.

3. Documentation variations

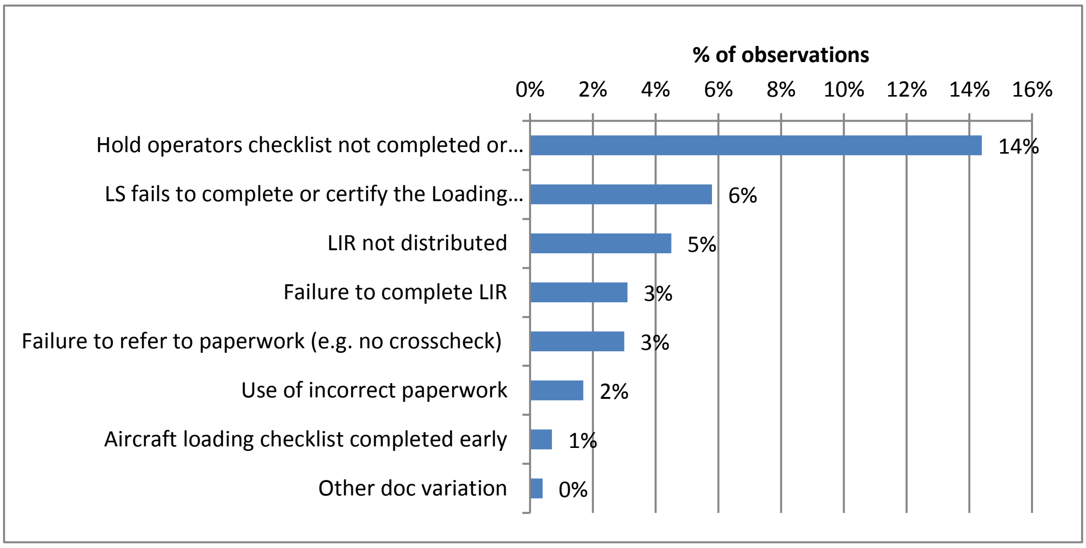


Figure 41 Percentage of each of Documentation variation subcategory occurring across all observations

There were 439 Documentation observations recorded across 330 observations. **Error! Reference source not found.** shows the most common type observed was the hold operator's checklist being incomplete or incorrect, affecting 14.4% of observations. This was followed by the Loading Supervisor failing to certify the loading checklist, affecting 6% of observations.

4. Restraining variations

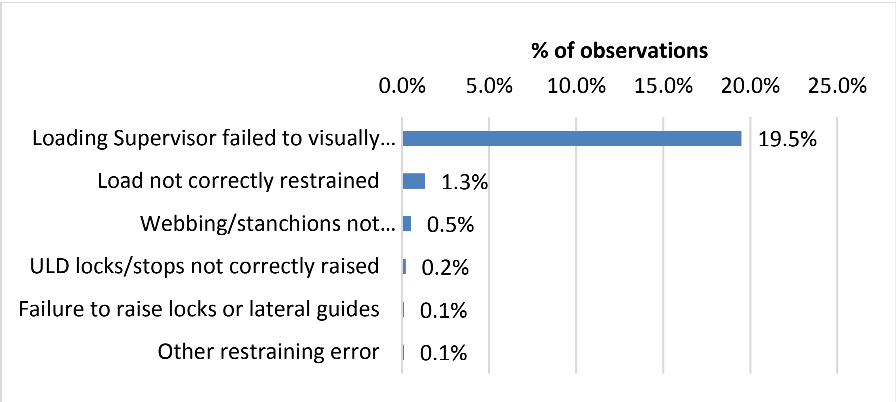


Figure 42 Percentage of each of Restraining variation subcategory occurring across all observations.

There were 289 restraining variations recorded across 270 observations. Figure 422 shows the most common type of restraining variation was the Loading Supervisor failing to visually inspect the doorsill locks and restraints at the end of the loading process, affecting 19.5% of observations. This was followed by the variation of the load being incorrectly or inadequately restrained, which was recorded in 1.3% of observations.

5. Readback variations

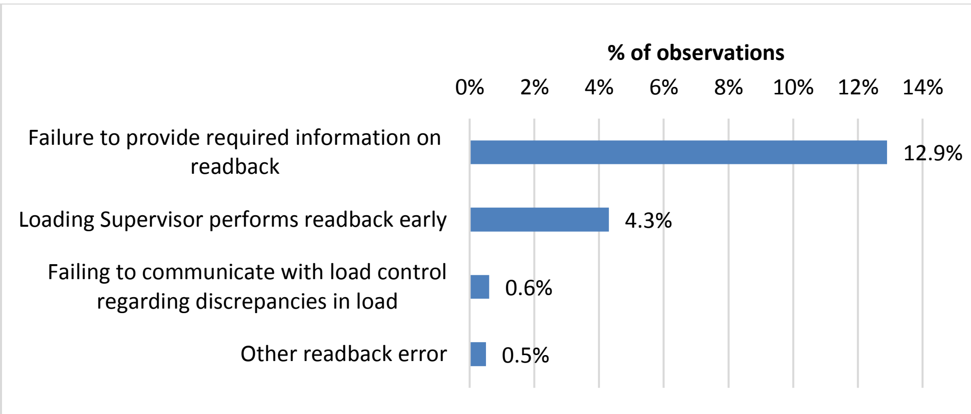


Figure 43 Percentage of each Readback variation subcategory occurring across all observations.

There were 246 Readback variations recorded, with 222 observations having at least one of this type. Figure 433 shows the most common was the Loading Supervisor failing to provide the required information when performing the read back, affecting 12.9% of observations. This was followed by the Loading Supervisor performing the read back early, recorded in 4.3% of observations.

6. Loading variations

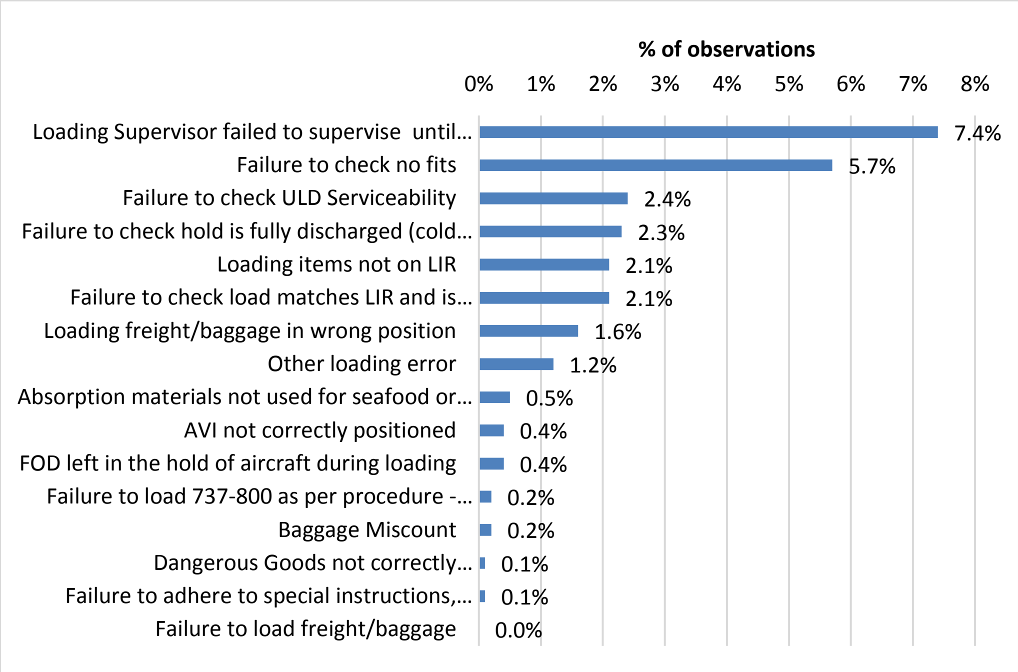


Figure 44 Percentage of each of Loading variation subcategory occurring across all observations.

There were 359 Loading variations recorded affecting 223 observations. Figure 444 shows the most common of these was the Loading Supervisor failing to supervise until all the load is onboard. This was recorded in 7.4 % of observations. This was followed by failing to check the hold for 'no fits' or empty positions, which occurred in 5.7% of observations. There were no occasions on which freight or baggage was left on the tarmac or failed to be loaded at all.

7. Unloading variations

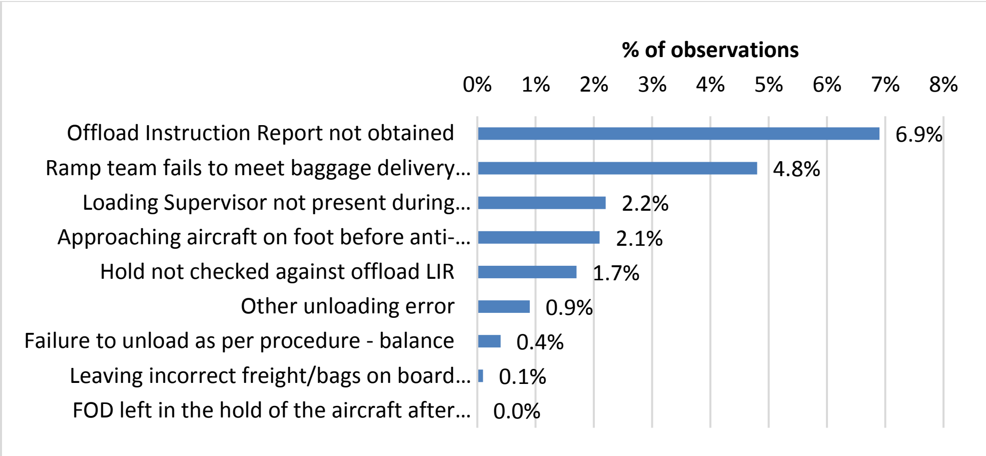


Figure 45 Percentage of each of Unloading variation subcategory occurring across all observations.

There were 251 Unloading variations affecting 197 observations. Figure 455 shows the most common subcategory was the failure to obtain an offload instruction report (or offload LIR) to check against the arriving load. This variation was recorded in 6.9% of observations. This was followed by the ramp team failing to meet baggage delivery standards regarding priority baggage, in 4.8% of observations. There were no occasions on which the ramp crew were observed leaving foreign object debris (FOD) during the unloading task.

8. Door variations

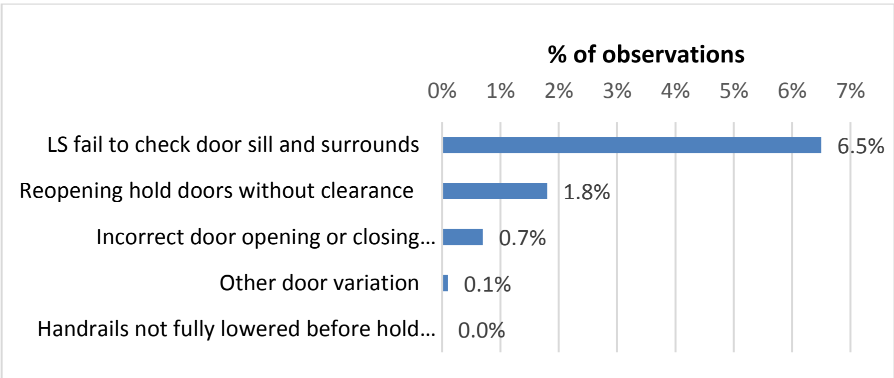


Figure 46 Percentage of each of Door variation subcategory occurring across all observations.

There were 119 Door variations recorded within 116 observations. Figure 466 shows the most common subcategory was the Loading Supervisor failing to perform a check on the doorsill and surrounds (6.5%), followed by reopening the hold doors without the proper clearance (1.8%).

9. Bay Management variations

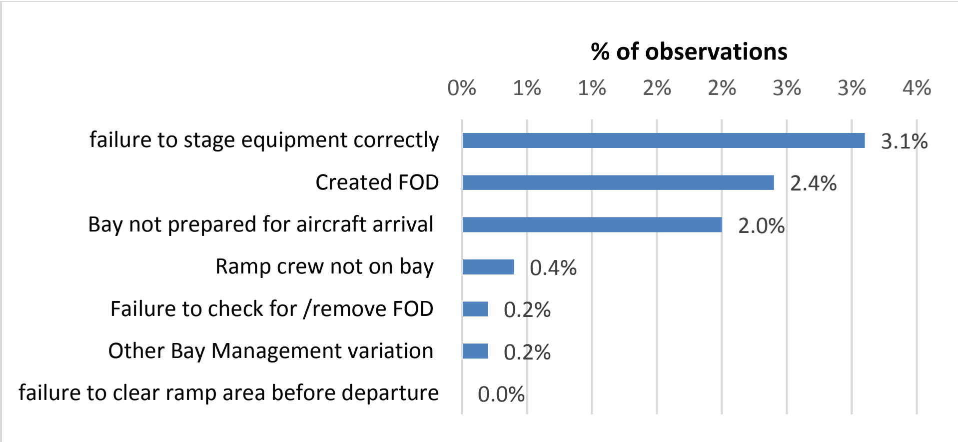


Figure 47 Percentage of each of Bay Management variation subcategory occurring across all observations.

There were 109 Bay Management variations with 102 observations recording a variation of this type. Figure 477 shows the most common subcategory was the ramp crew failing to stage equipment correctly in the marked areas (3.1%), followed by the ramp crew leaving foreign object debris (FOD) on the ramp, recorded in 2.4% of observations.

10. Communication variations

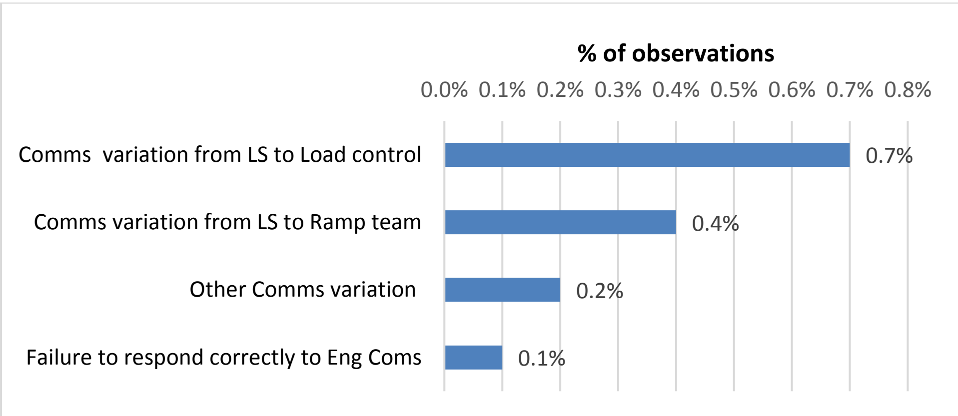


Figure 48 Percentage of each of Communication variation subcategory occurring across all observations.

There were 17 observations recording a Communication variation. Figure 488 shows the most common was a variation in communication from the Loading Supervisor to Load Control, recorded in 0.7% of observations. This was followed by communication variations between the loading supervisors and members of the ramp team, noted in 0.4% of observations.

9.3.4 Most frequent subcategory variations

Some variations were observed more commonly than others. The five most frequently occurring variation subcategories are shown in Table 17. Examples of each variation type are provided in the descriptions below the table.

Table 16: Five most frequently occurring variations.

Variation types	Major Threat Category	Total obs. with this variation	% obs. affected
1. Failure to perform circle of safety	Driving Variation	573	44%
2. Leaving motorised vehicle unattended	Driving Variation	564	43.3%
3. Control pendant not placed in hold	Equipment usage Variation	375	28.8%
4. Loading supervisor failed to visually check restraints	Restraining Variation	254	19.5%
5. Failure to perform pre-driving inspection	Driving Variation	212	16.3%

1. **Failure to perform circle of safety (ramp crew).**The most commonly occurring variation was a driving variation where the ramp crew failed to perform the circle of safety. This occurred in 44% of observations. Example from narrative:

> *'It was noted that the operator of the forward machine did not conduct the circle of safety as no second stop was applied before the machine was positioned to the aircraft'.*

2. **Leaving motorised vehicle unattended.** The second most common variation was also a Driving variation, where vehicles were left running but unattended near the aircraft. This occurred in 43.3% of observations. Example from narrative:

> *'There were six crew members on the bay, including the Loading Supervisor, who was observed walking to the belt loader which had been left running in the equipment storage area. The Loading Supervisor then switched off the engine which had been left running'.*

3. Control pendant not placed in hold. The third most common variation was an Equipment usage variation, where the crew failed to put the control pendant in the hold when the belt loader is in position. This occurred in 28.8% of observations. Example from narrative:

> *'As the aircraft was being unloaded it was identified that the control pendant at the forward end of the belt loader at the AFT hold had not been positioned inside the hold of the aircraft'.*

4. Loading supervisor failed to visually check restraints. The fourth most common variation was a Restraining variation, where the Loading Supervisor fails to check the correct restraints are in place to prevent the load from moving during flight, occurring in 19.5 % of observations. Example from narrative:

> *'On completion of the read back the bulk hold door had been secured and the Loading Supervisor had not inspected the webbing prior to the door being closed'.*

5. Failure to perform pre-driving inspection. The fifth most common variation was a Driving variation that involved drivers of ground servicing equipment failing to perform a pre-driving inspection. This occurred in 16.3 % of observations. Example from narrative:

> *'Once in position it was identified that a piece of plastic was hanging from the bridge of the JCPL. The plastic was not removed and was approximately 1m long'.*

9.3.5 Where variations occurred

There were more than twice as many variations at international ports as at major domestic or regional domestic ports.

Table 17: Mean variation per observation for each port type.

Port Types	Number of obs	Mean	Standard Deviation
International	298	5.73	3.012
Domestic	820	4.00	2.333
Regional domestic	73	2.70	1.543

The mean number of variations recorded during any one observation was significantly higher at international ports (SD 5.73) as shown in Table 17

Other possible analysis of variations

The data collected regarding variations allows for analysis in a number of different ways, For example, in terms of port, airport bay, aircraft type or team size. It is also possible to use information from narratives, observer issues or port reports, to provide additional context to the variations occurring. Examples of how this data might be further analysed are provided below.

Example analysis of port reports

Port reports allow the observers to enter context about local conditions in each port that may help to form hypotheses about why variations may be occurring in that location, which can then be further investigated. The following examples from port reports provide additional context to livestock loading variations and Loading Supervisors completing the read back early:

Livestock loading variations. *'[Port C] s has its own local procedure regarding the loading of AVI (livestock) and this procedure is not documented. When AVI is loaded and secured in position 11 on the B737 aircraft, the fwd. cargo door is left open for the purpose of ventilation for the AVI. The fwd. cargo door is closed by the engineer prior to departure. This may be an issue as the Loading Supervisor will have completed the read back and the Captain will receive the final load sheet whilst the fwd. cargo door is in the open position'*.

Completing read back early. *'According to the Precision Time Schedule rules in [Port H], the Loading Supervisor must complete the read back to greater than 5 minutes in order to give load control time to update ACARS. If you're late with the read back (even if the Captain gets the Load sheet of ACARS in time) passenger services can pass a delay onto the ramp. This practice may be encouraging the loading supervisor to complete the read back early before the final load is on board'.*

Analysis of observer issues.

Another example of further investigation that can contribute to the understanding of variations, is the analysis of observer issues. Observers add additional information that may be relevant to variations in the 'observer issues' field in the database. Observer issues provided context about procedures that are ambiguous, difficult to comply with, or had been implemented differently in some locations. Examples of observer issues relevant to the coding of variations are provided below:

Procedures regarding the circle of safety. *'The ramp manual includes a diagram with a colour coding mistake. The diagram shows the areas where Ramp Team members can and cannot drive within an aircraft's circle of safety. This text referring to the diagram states that the location of the "vehicle no-go zones", are shown in black "vehicle no parking/no standing zones (fuel venting points)" shown in yellow, and the "slide deployment zones" are shown in red. However in the diagram "slide deployment" zones are yellow and the fuel venting points red. This may be causing confusion'.*

Procedures for the loading of animals. '*The opening of cargo doors for animals in port A is an unofficial solution to the problem of managing two conflicting rules: Firstly that the Loading supervisor must be responsible for closing cargo hold doors and that they cannot be reopened without the appropriate clearance. Secondly that there is a need to ensure livestock have appropriate ventilation*'.

Exemptions for driving under the wing. '*The procedure states that a GSE operator must never drive underneath the fuselage or wing(s) of an aircraft unless operationally required for servicing however there appear to be approved local practices for driving under the fuselage of aircraft in some airports. The local instructions say operators must familiarise themselves with the approved practices applicable at their ports which may be causing some confusion'.*

Procedures for leaving vehicles unattended. *'The procedure states that GSE operators must never leave vehicle engines running while the vehicle is left unattended. It is not clear whether this includes dismounting from the vehicles to attach equipment to the back, or whether it is necessary to switch off the engine every time. In addition some ports have older GSE which requires jump starting. When this occurs it is common for equipment to be left running for a period of time in order to charge the battery. It is not clear how the rule applies in this case'.*

9.4 Analysis of Undesired Operational States (UOS)

9.4.1 Overview of Undesirable Operating States (UOS)

UOS are situations where safety may be compromised as defined in Chapter 8. There was a total of 165 Undesired Operational States, or UOSs, recorded in the observation period. 102 were associated with at least one variation, and 41 were associated with at least one threat. There were 22 UOSs where no preceding threat or variation was associated. Table 18 shows the total UOSs in each category and provides examples of each from the narratives.

Table 18: the number of UOS occurring in each subcategory with examples.

Total UOS per major category	UOS subcategory	No. UOS per sub-category	Examples of each USO from narratives
Loading UOS Total : 67	Irregular load other	31	*The freight satchel, papers and crew bags were then loaded into the bulk hold of the aircraft. These items were placed into position 51. The freight satchel and papers were not listed on the LIR. This was not communicated to Load Control during the Read back.*
	Loading of unserviceable ULD	16	*It was noted that AKE 37546 had a hole in the panel alongside the frame work approximately 2centimeters long. As this was on the edge of the frame work it was also within 5 centimetres. This was not identified by the loading supervisor before it was loaded.*
	Load not in the correct position	8	*Loading Supervisor inspected the freight and was observed cross checking the freight against his Load Instruction Report (LIR). He collected and removed all the barrow cards from the barrows but did not identify the box which was now on the incorrect barrow. The box was eventually loaded into position 31 of the aircraft when it should have been loaded into position 21.*
	Unrestrained load	4	*The load in position 9 was completely unrestrained when the hold door were closed.*
	Weight and balance discrepancy	4	*The ULD card on AKE23314 was 1kg lighter than the weight indicated on the LIR and this was not identified by the Loading Supervisor.*

Total UOS per major category (cont'd)	UOS subcategory (cont'd)	No. UOS per sub-category (cont'd)	Examples of each USO from narratives (cont'd)
Loading UOS (cont'd)	Freight weight discrepancy	2	*The loading supervisor approached the forward cargo door; he opened it and looked inside... The LIR showed an AVI and a small amount of freight in hold 11, and bags in hold 21. The loads in the holds did not match the LIR; this was an irregular load.*
	Incorrect baggage or freight loaded	1	*During the final read back the Loading Supervisor read these containers being in opposite positions so that AKE26308 was in 15R and AKE23967 13R. These were in the doorway of the FWD hold and not changed before the cargo door was closed.*
	Other loading UOS	1	*2038 A yellow documentation bag was placed on the belt loader to be loaded into the aircraft bulk hold it was noted that this bag was not documented on the LIR. This was not mentioned to load control via the read back.*
Conflict / Near Collision equipment to aircraft UOS Total:25	Conflict/Near Collision with engines	9	*The belt loader is removed from the aircraft; the hand rail is still in the raised position. While it is being driven away and narrowly misses a collision with the engine.*
	Conflict/Near Collision with fuselage	7	*The engineer notices that the cargo handling panel has been left open, he informs the Loading Supervisor The operator drives to the forward hold quickly, he fails to perform the circle of safety and does not slow down until close to the aircraft, he sharply applies the brakes and the JCPL stops with the bridge under the fuselage, the safety rails are within 50mm of the fuselage. The Loading Supervisor sees this and looks over at me and tells me that it was very close.*
	Other conflict/near collision of equipment to aircraft	4	*It was also noted at this time, that the forward platform on the catering truck at the R1 door was hard up against the aircraft fuselage. It was positioned with enough pressure that the buffer on the platform was half squashed.*
	Conflict/Near Collision with doors/sills	4	*The first container in the doorway of the forward hold is removed, the operator is not watching the container, as it comes out of the doorway the ULD spins to the right and an impact with the door sill, the aluminium is flattened and a proper seal with the door rubber will now not be possible. The operator does not inform anyone of this and continues unloading.*
	Conflict/Near Collision with wings	1	*The belt loader at the front [of the aircraft] has defective brakes as the staff was placing A chock behind the rear wheel so as to stop the belt rolling back. [However] they did not place a chock in front of the wheel so the belt loader was able to move forward and damage the aircraft.*
Delay UOS Total: 23	Delay event	23	*No ramp team was available for unloading or loading of the aircraft leading to 30 min delay.*
Livestock UOS Total :20	Unsecured livestock	19	*Three AVIs are delivered from cargo to the aircraft and are loaded immediately. The LIR [correctly planned the] AVI into position as per the [standard operating] procedure. There are no tie down restraints supplied by cargo so the animals are left unsecured.*
	Other livestock issues	1	*It was observed that an absorbent material was placed under the AVI cage and restrained with rope.*

Total UOS per major category (cont'd)	UOS subcategory (cont'd)	No. UOS per sub-category (cont'd)	Examples of each USO from narratives (cont'd)
Conflict near collision between equipment UOS Total: 16	Vehicle conflict /near collision with fixtures	6	*The driver of the JCPL narrowly missed the side of the terminal building while turning.*
	Vehicle conflict/collision with other vehicles	5	*The driver returned to the tug towing a train of four dollies and positioned to the rear of the AFT JCPL. In an attempt to get as close as possible to the rear of the loader the first dollie struck the rear of the JCPL. As each ULD was unloaded and the driver re-positioned the dollies each dollie* [sic] *contacted with the rear of the JCPL.*
	Other conflict/near collision with equipment	3	*A driver positioned another low profile to the rear of the FWD JCPL. The driver reversed the low profile and there was an operator at the bottom of the JCPL but did not marshal the driver. The low profile then hit the rear of the machine.*
	Falling object	2	*A drum has been loaded on the baggage chute… the drum rolled down the chute and collided with the push out vehicle.*
Ground Event UOS Total:9	Other ground event	5	*A Ramp Supervisor is on the bay he is on his mobile phone whilst being within 2m of the Refuelling truck.*
	Jet blast/ jet intake	2	*An A380 which had been on the adjoining bay was in the stages of departure and was just leaving the bay. As the aircraft taxied past the bay and turned towards to the runway the ramp team on the bay were hit with jet blast. This created a lot of FOD which was now blowing around the aircraft. The first AKE was then loaded onto the aircraft.*
	FOD- potential for FOD damage	1	*As the aircraft taxied onto the bay a piece of plastic strapping that was lying on the taxi way got sucked into the aircraft's engine.*
	Speeding within footprint of aircraft being serviced	1	*Once the path was clear the driver climbed into the pushback tug and moved it into position for the pushback. The tug was driven at excessive speed around the aircraft as big clouds of black smoke could be seen coming from the exhaust of the tug and the motor was revving loudly. When the tug came to a stop excessive braking was required and caused the tug to rock and come into contact with the nose wheel.*
Injury /Near miss to personnel Total: 3	Injury/near miss with personnel	3*	*The aircraft arrived on the bay and was approaching the hold line when the Engineer opened the communications flap on the aircraft. The aircraft was still moving while the Engineer connected the wireless headset receiver.*
Dangerous Goods UOS Total : 2	Unrestrained dangerous good	2	*The box was an 80kg shipment containing Dangerous Goods which was also observed being secured. The item was secured with rope and not a company approved restraint.*

*One UOS resulted in a recordable incident; in this incident the locks were not raised on a low-profile ground servicing equipment vehicles and Unit Load Device fell off as the operator tried to move the train of rolling stock, which resulted in a personnel injury.

Loading UOS were the most commonly occurring major UOS category; the largest UOS type in this category was 'other' irregular load (n31) which can cover an eventuality where the load does not match the load instruction report. There were 16 incidents where unserviceable or damaged Unit Load Devices (ULD) were loaded on to the aircraft with the potential to cause damage to the hold or fuselage. The next most frequently occurring major category for UOS was Conflict or Near Collision of equipment to aircraft, most typically contact or near contact with aircraft engines (n9) and the fuselage (n7). There were 23 occasions where a severe delay was recorded as an undesired operational state.

9.4.2 Where Undesired Operational States occurred

Table 19: Number of Undesired Operational States occurring at International and Domestic Ports.

Port Type	Number of UOS	Percentage of UOS
International Ports	94	57%
Major Domestic Ports	69	41%
Regional Ports	2	1.2%

Undesired Operational States were more common at international ports (57%) than large domestic ports (41%) and were very rare at minor domestic regional ports (1.2%), as shown in Table 19. UOSs can be analysed in a number of different ways, not all of which can be presented here. For example, Undesired Operational States can be analysed by the percentage of turnarounds involving each aircraft type.

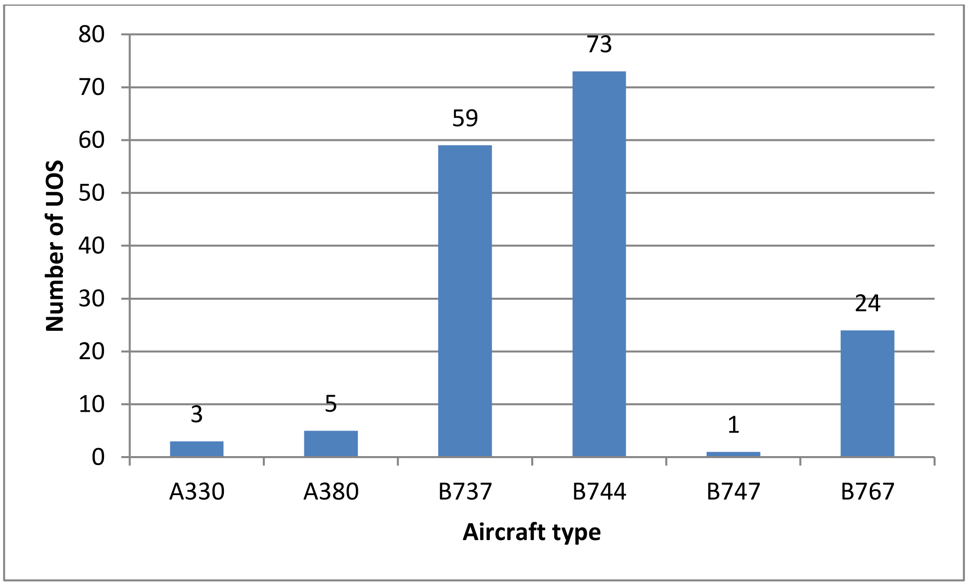


Figure 49 Number of Undesired Operational States occurring for each aircraft type.

Figure 4949 indicates that many of the Undesired Operational States occurred on the B744 aircraft type (n73) with only one recorded for the B747 aircraft type. When compared with the number of observations conducted, we can see that the B744's UOS rate is high considering it only represented 11% of all aircraft observers, as shown in table 21 below.

Table 20: Number of UOS by aircraft type and % of observations.

Aircraft Type	B737	B767	**B744**	A330	B17	A380	D8	A320	B747	B777
No. of obs.	769	236	**144**	56	37	26	16	8	8	2
% of total obs.	59.06	18.12	**11.05**	4.30	2.84	1.99	1.22	0.61	0.61	0.15
Number of UOS	59	24	**73**	3	0	5	0	0	1	0
% obs for aircraft type	8%	10%	**51%**	5%	0	9%	0	0	13%	0

This can be investigated further to understand, for example, what kind of undesired states are most likely to occur on the B744 fleet.

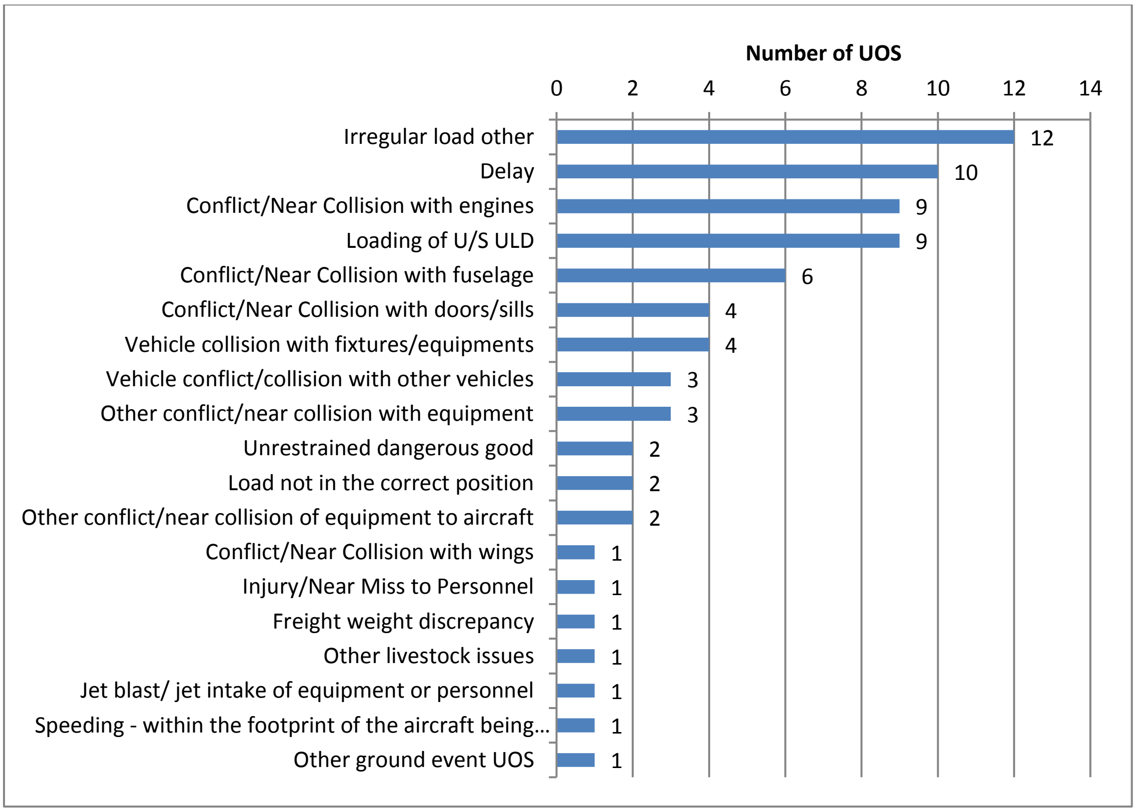


Figure 50 Number and type of UOSs occurring on B744 aircraft.

Figure 50 shows that the most common UOS for the B744 aircraft was 'other irregular loads' (n12) followed by conflict or near collision with the aircraft's engines (n10). Narratives can be used to provide further context of undesired states occurring on each aircraft type, such as this example of a UOS on a B744:

'The catering truck has completed loading the fwd. galley at door 1R. The driver lowered the truck then is marshalled away from the aircraft. The marshaller then marshals the truck under the wing area of the aircraft and very close to number 4 engine to get around the bay congestion of rolling stock positioned between bay 41 and bay 43'.

The UOS types occurring on this aircraft are different to the pattern observed overall, with higher numbers of conflict/collision UOS. It may be that this fleet is more likely to be deployed to busy overseas port with high congestion threats. Differences such as these could be investigated further to understand why, for example, the B744 is over-represented in the UOS data.

Further analysis of UOSs could also consider analysis by geographic location, aircraft bay or crew size, for example, with further context provided from observer issues analysis or port reports. Further discussion of the relationship between threats, variations and undesired states is provided in Chapter 10.

9.5 Analysis and Comparison of Management Rates

Observers recorded how often the threats, variations and Undesired Operational States were managed, mismanaged or unmanaged according to the definitions outlined in Chapter 8. This section provides an overview of how threats and variations were managed by ramp crew during this NOM implementation, and compares management rates with other LOSA style programs where data is available. Narrative examples are provided to add additional context.

9.5.1 Management rates for threats

Table 21: Managed, mismanaged or unmanaged threats.

	Number of threats	% of all threats
Managed	347	11.9%
Mismanaged	76	2.6%
Unmanaged	2500	85.5%
Total number of threats = 2923		

The majority of threats in this observation period were coded as ‘unmanaged’, with no action observed in response to the threat. Table 21 shows that only 11.9 % of threats were reported as managed, 2.6% were mismanaged, with the majority, 85.5%, recorded as unmanaged. Examples of managed threats are not common but are of interest to understand the circumstances in which management actions do occur.

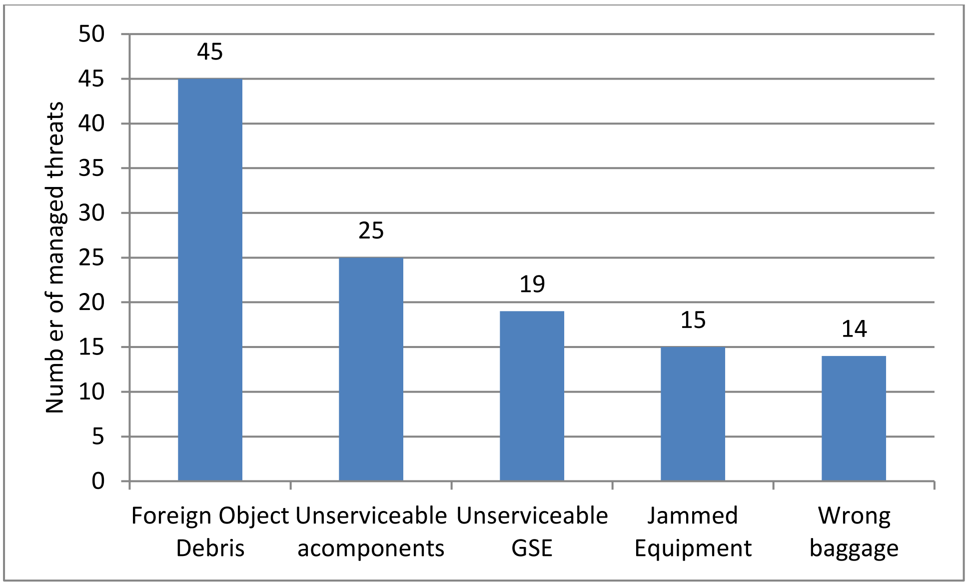


Figure 51 Threats most likely to be managed.

Figure **51**51 shows the five threats with the highest rates of managed behaviours recorded. The threat of Foreign Object Debris was managed on 45 occasions, followed by unserviceable aircraft components, which were managed on 25 occasions. Narratives can be examined for behaviours that were successful for managing threats, as they may provide useful context for considering interventions or as training case studies.

Managed FOD threat. ‘*On visual inspection of the aircraft bulk hold there was a significant amount of FOD inside the hold. This consisted of ten pieces of timber and some plastic bottles that were filled with water. There was also a large piece of plastic and four company brown lines. The LS supervised the removal of the FOD before commencing the loading processes’.*

9.5.2 Management rates for variations

Table 22: Managed, mismanaged and unmanaged variations.

	Number of variations	% of all variations
Managed	71	1.4%
Mismanaged	72	1.4%
Unmanaged	5047	97.2%
Total number of variations = 5190		

Almost all the variations observed were coded as 'unmanaged', meaning no action was taken in response to the variation (97%). Correspondingly low numbers were managed (1.4%) and mismanaged (1.4%) respectively. Examples of managed variations were very rare but of interest as they can highlight circumstances in which management actions do occur.

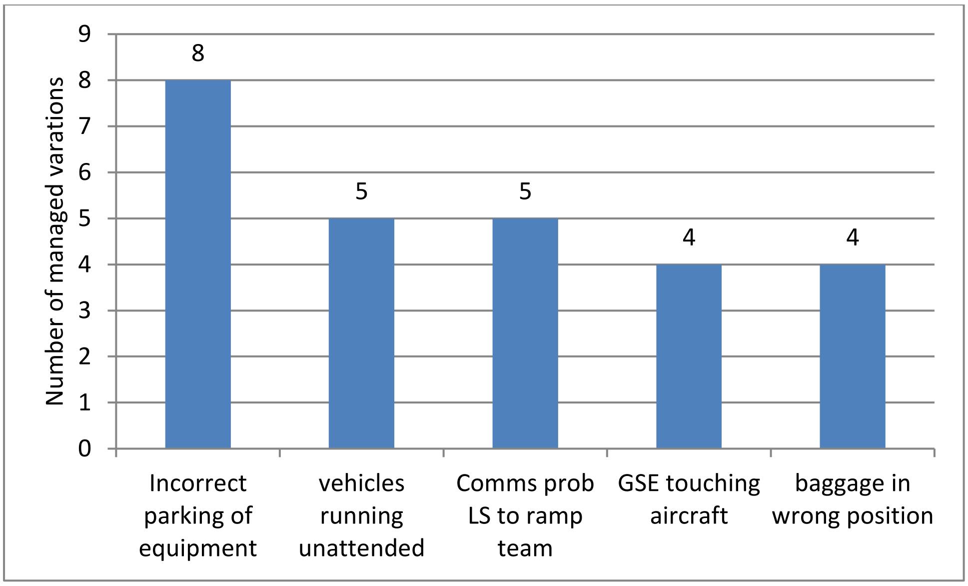


Figure 52 Variations most likely to be managed.

Figure 522 shows the variations with the highest management behaviours recorded. There were eight occasions on which the variation of incorrect parking was managed, five occasions on which vehicles left running unattended near the aircraft were managed, and five examples where communication issues between the Loading Supervisor and the ramp team were managed. Narratives provide useful examples for training case studies.

Managed variation: freight and baggage in wrong position. *''The first container AKE26308 was loaded into 13R of the FWD hold and the second ULD AKE23967 was loaded into position 15R. During the final read back the Loading Supervisor realised the containers were in opposite positions so that AKE26308 was in 15R and AKE23967 13R. These were changed and then the cargo door was closed'.*

9.5.3 Management rates for Undesirable Operational States

Table 23: Managed, mismanaged and unmanaged Undesirable Operational States.

	Number of UOS	**% of all UOS**
Managed	10	6.1%
Mismanaged	3	18%
Unmanaged	152	92.1%
Total Undesired Operational States = 165		

Table 23 shows that while UOSs were more likely to be managed than variations, the majority of Undesired Operational States (92%) still go unmanaged, whilst 18% are mismanaged and only 6.1% are managed. Examples of managed UOSs were extremely rare in this dataset.

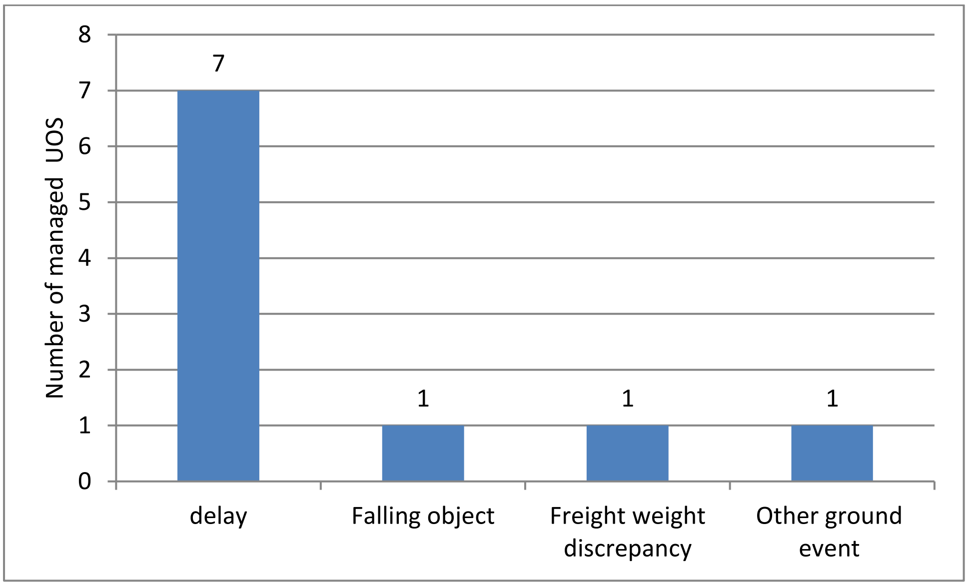


Figure 53. Comparison of management rates across major categories.

There were only 10 examples of managed UOSs. Seven related to managing severe delays, a falling object, one freight and weight discrepancy and one managed ‘other’ ground event. Examples from narratives provide further context to understand how UOSs were managed.

Managed UOS: falling object. ’*the train of three dollies is towed forward from the AFT JCPL the driver turns towards the wingtip and the second low profile’s locks are not raised and PMC 42632 falls off the low profile and the driver stops when he hears the PMC hit the ground, with the driver stopping it keeps the whole PMC from settling on to the ramp and it is supported by the low profile at about a 45 degree angle, the leading hand uses his radio to locate a forklift driver’.*

9.5.4 Management rates compared with other LOSA-style programs

One of the participating airlines had existing observation programs based on the LOSA methodology and agreed to make this data available for the purposes of comparison with NOM ramp data. When compared with data from other LOSA style programs within this airline, the management rates for threats, variations and undesired states for the ramp NOM implementation were comparatively Low. Although there are differences in the NOM and LOSA methodologies, definitions for management of threats, errors (variations) and UOSs remained similar, making for interesting comparisons of management rates across different types of programs.

Figure 53 shows the ramp NOM data shows considerably lower threat management rates (11.9%) when compared to the percentage of managed threats in the flight crew LOSA program (92%) and cabin LOSA program (76%). This data must be viewed with caution, however, because the differences in the working environment and training of crew have not been accounted for.

Similarly, the management rates for variations (referred to as errors in the LOSA program) are also low by comparison with this airline’s internal programs, as shown in Figure

534. Only 1.4% of NOM ramp variations are reported as managed, compared with 38% in the flight crew LOSA program and 19% in this airline's internal LOSA program. Again, the data must be viewed with caution as differences could be a function of the differences in NOM and LOSA methodology.

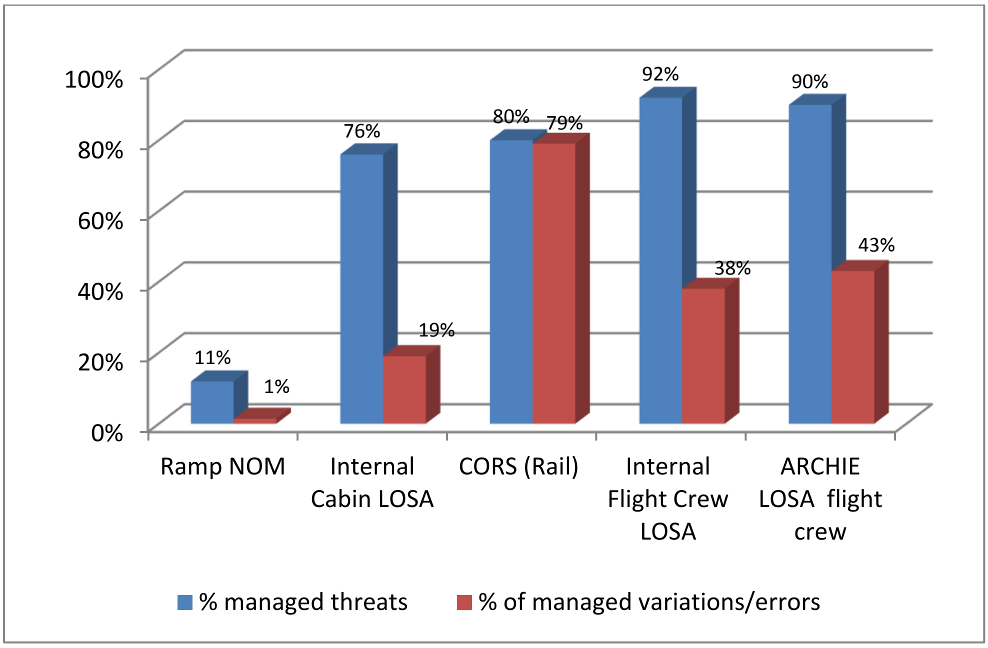


Figure 53: Comparison of management rates across programs.

Although there are obvious environmental and methodological differences between programs, it is nevertheless interesting to investigate differences in management rates from other LOSA programs. Data has been published from both the CORS Rail study (McDonald and Barragan 2007) and the LOSA Collaborative's combined ARCHIE database of LOSA data from flight operations (Merritt and Klinect 2007). It is, therefore, possible to compare management rates between these programs.

Figure 534 shows that ramp NOM threat management rates are low (11.9%) when compared against the rail CORS data, which has a threat management rate of 80%, and the Flight Operations ARCHIE data for LOSA, where 90% of all threats are managed.

9.6 Discussion

One objective of the study was to collect a large sample of data that would be representative of normal ground operations. The dataset suggests that this objective was achieved, with over 1300 observations conducted across six different carriers' global operations. The observations cover a broad range of locations, taking place in 28 ports across 10 different countries. They included 10 different aircraft types observed from a range of positions around the aircraft, with crew sizes from 2 to 12 team members. Therefore, it is possible to be reasonably confident that the data has captured a reasonable sample of 'normal operations' from the ramp environment.

The data collected in each observation includes both qualitative and quantitative data. The hand-written narratives, for example, provided a rich data set from which insights about the context of the turnaround could be explored. The coding of the data in terms of the threats, variations and undesired states provided many novel insights into the nature and extent of issues occurring that are affecting ramp safety outcomes.

9.6.1 Distribution of threats in ground operations

The analysis of threats highlighted a number of recurring themes in the data. For example, airside driving is a particular problem across all teams in all locations.

Airside Driving. Airside driving threat data highlights the poor driving practices of non-ramp crew, which present as a threat to ramp safety. Driving variations indicate poor driving of the ramp crew themselves. It would appear that most groups who drive on the ramp were observed engaging in poor driving practices. The most frequently occurring threat was the failure to follow the circle of safety (by non-ramp crew); the low compliance rates with this procedure suggest it is ineffective as a control for preventing conflicts and collisions with aircraft. In nearly half of all observations, this procedure was not performed. Further consideration should be given as to how this procedure could be amended to increase compliance, if appropriate, or whether the risks of driving in close proximity to aircraft could be better controlled another way.

Equipment usage threats. Equipment usage threats were common. The most frequent included GSE being incorrectly parked around the aircraft and the incorrect use of refuelling cones to prevent GSE incursions around fuel lines. There were also several related threats where the ramp crew faced equipment issues, such as unserviceable, damaged or jammed equipment, or when no equipment was suitable for the task.

Congestion. The staging of equipment and associated congestion appeared to be an issue affecting both airside driving and bay management. Failure to park ground servicing equipment in the designated areas causes obstructions and forces others to drive under the wing of an aircraft in order to move around the bay. Port reports revealed a general lack of parking at airports, particularly at large busy ports where parking areas were insufficient for the amount of equipment. In some ports there were no designated areas for equipment, which meant that staff parked GSE wherever it could fit. This then caused the area to be congested and limited the movements of the rolling stock and equipment on the bay. The analysis of threats indicates that while some threats such as parking may not seem particularly high risk, the flow-on effects may significant, as the behaviour of crews adapts in response to the threat by diverting around obstacles and finding paths through prohibited areas (such as under aircraft wings). With airport traffic and congestion expected to increase in the next five years (IATA 2014), parking and storage of equipment appears to be an issue that requires further analysis, if we are to avoid some of the driving violations that lead to aircraft damage.

There also seemed to be issues related to the serviceability of equipment, such as unserviceable components on the aircraft themselves or equipment jamming or becoming unserviceable during use. Where the right equipment is not available or is unusable, crews are forced to improvise. The risks of not using the right equipment need to be assessed. The data could be further investigated to explore where new equipment is most needed to help allocate resources effectively.

Foreign Object Debris (FOD). FOD being left on the ramp was third largest threat category. FOD can vary in form from relatively harmless scraps of paper to more substantial objects that can be ingested into engines. The majority of the FOD threats recorded were of FOD left on the ramp, rather than in the aircraft hold. FOD threats on airport runways have caused major aircraft accidents, such as the Concorde Flight 4590 accident in Paris in 2000, which killed 113 people. In this case, foreign object debris on the runway blew a tyre, puncturing a fuel tank and leading to fire and engine failure (Bureau Enquêtes-Accidents 2000). The National Transport Safety Board (2007) has also highlighted the contribution of FOD to serious accidents in aviation. Further investigation is needed to understand the sources of FOD and consider ways in which it can be further eliminated or controlled.

Threats from other airport teams. Most of the threats recorded during the observations came from the ramp team's interactions with other ramp crew. The ramp team interacts with many other teams whilst performing the turnaround, including catering, engineering, Load Control etc. Interfaces refer to the activities on the ramp where ramp tasks interact with other segments or third parties either directly or indirectly. All of these interfaces have the opportunity to increase threats and errors to the loading process. The analysis of the origin of threats highlights the impact that different airport crew have on the ramp operation. In this dataset, it is possible to see that threats most commonly arise from other ramp crew servicing nearby aircraft. As might be expected, the second largest group of threats arise from catering activities, followed by Freight and Baggage services. This information allows the organisation to target interactions that impact safety on the ramp.

Freight and Baggage threats. Threat data also revealed that issues arising from the Freight and Baggage areas were particularly problematic for crews during the turnaround window. The most commonly occurring problems were incorrectly profiled freight, such as incorrect labelling or stacking of loads, as well as loads that were incorrectly restrained. Incorrect labelling can occur, for example, when old baggage cards are left on containers. This common threat has the potential to cause confusion for ramp crew who cannot reconcile old labels with current documentation. Observer issues revealed there were no documented procedures outlining who has the responsibility to remove old barrow cards, with procedures varying from port to port. In some ports there is a belief that the barrow cards should be removed by the last personnel to handle the barrows. Confusion regarding this procedure may be contributing to the high frequency of this threats observed.

Freight and Baggage threats are likely to increase the workload of ramp crews, who find themselves trying to resolve problems during the turnaround and divert the attention of the Loading Supervisor away from other supervisory tasks such as checking restraints and locks, for example.

Due to the compressed timeframes, ramp crew have very little time to manage threats during the turnaround, suggesting that these threats would be most effectively managed at source in the freight area, for example, or introducing new guidelines for ramp crew to reject freight that is not correctly presented.

9.6.2 Distribution of variations in ground operations

Airside Driving. Given the prevalence of poor driving practices observed on the ramp, it is not surprising that airside driving was also the most prevalent variation for the ramp crews under observation. Driving variations for ramp crew were very similar to the threats observed in non-ramp crew, such as failing to observe the circle of safety. Observers recorded this as a variation when a ramp team member failed to perform this procedure. Comments from observers revealed some ambiguity in the colour coding of the diagram used to indicate the circle of safety zones around an aircraft in the ramp procedures manual, but this is probably not sufficient to explain the large number of variations from this procedure. Non-compliance with the circle of safety was high for all drivers accessing the ramp. The very low compliance rates (less than 50%) indicate that the feasibility of the circle of safety procedure needs to be reviewed.

The observations also revealed other variations with the potential to result in aircraft damage – for example, leaving vehicles unattended or not using a parking brake, both of which have the potential to result in GSE rolling into the aircraft. The ramp procedures state that GSE operators must never leave vehicle engines running while the vehicle is left unattended. Observer comments suggested that there may be some interpretation as to what constitutes 'leaving the vehicle', such as when attaching something to the back or after jump-starting older equipment to charge the battery; these issues make the rule difficult to comply with in practice, which may be contributing to some of the variations observed. Towing more than permissible amounts of rolling stock may also increase the risk of aircraft damage. The extra weight can affect the braking performance of GSE operating around the aircraft and result in collisions. Port reports indicate that there may be differences in the allowable rolling stock at different locations.

Problems with procedures. The analysis of these variations suggests that, in some cases, the rules themselves are ambiguous or difficult to follow. Impractical rules are likely to increase noncompliance and discredit rules generally as a control for risks. High variation rates therefore suggest that these rules and procedures should be targets for review and reform. If procedures cannot effectively control the risk, then alternative controls may be necessary, such as speed limiting devices on vehicles that have been introduced in some airports (Bishop et al. 2008) and surface detection systems for airports (Bass 1995).

Equipment variations. Like equipment threats, equipment variations also overlapped with driving variations, as much of the equipment is also driven. Looking at driving and equipment variations together also highlights the potential for interactions to increase the risk of accidents. For example, driving under the wing may not be a problem on its own, but when combined with the equipment variation of leaving handrails or batwings extended, contact with the aircraft becomes more likely. The interaction between threats and errors is explored further in Chapter 10.

Congestion in busy ports. Variations were twice as common in international ports. This is not unexpected, as international ports are typically very busy and congested environments. This means there may be fewer approved driving routes and smaller staging areas for parking GSE safely, for example. Crew are then forced to vary the procedures with regard to driving close to aircraft or structures or around obstructions. It is noted that large international ports have seen an increase in the number of flights handled every year and this trend is likely to continue (IATA 2014), which reinforces the need to review the effect congestion may have on the safety of airports in the future.

Rules and rule compliance. The large number of variations suggests where rule compliance is low. In some cases, this may be due to differences in local procedures or problems with the rules themselves, such as the different interpretations of the procedures for the circle of safety limits, driving under the wing or leaving vehicles unattended, as discussed in 9.3.4. The NOM methodology records all deviations from standard operating procedures (SOPs) as variations, but this is not to assume that the SOPs are necessarily correct. High levels of non-compliance demonstrated with NOM data variation indicate where intervention is necessary to either improve the rule or improve compliance.

9.6.3 Distribution of Undesired Operational States in ground operations

It was anticipated that Undesired Operational States, or UOSs, would be rare in normal operations. It was therefore surprising that there were 165 UOSs recorded during the observation period. If the data is representative, this would mean that UOSs are occurring in over 11% of all normal operations, which is concerning. The most common UOS types were loading events where the actual load did not match the load as planned and documented. This can affect the weight and balance of the aircraft and is also a regulatory breach that can lead to fines (n67). Undesired Operational States resulting in conflicts or near collision with aircraft were the second highest category (n25). Given the urgent need for the industry to address the cost and injuries associated with such ground damage incidents, as discussed in Chapter 2, preventing these events would be a key area to target for intervention.

UOS are more common in international ports and on certain types of aircraft typically deployed to busy international ports. Further investigation is required to investigate why this might be the case.

Consideration of the risks associated with each UOS would be a useful area for future study. Putting additional controls in place for high risk UOSs, as well as targeting the threats and variations that precede them, should therefore be a key area of focus to prevent their reoccurrence. The associations between threats, variations and UOSs are discussed further in Chapter 10.

9.6.4 Management rates

Overall, the management rates of threats and errors and undesired states on the ramp in this study were particularly low. When compared to LOSA studies, the results of this research suggest that ramp crew very rarely take action to manage threats and variations. There may be several reasons for this, which should be explored in future research. When compared to flight crew, for example, ramp crew receive little or no training on how to detect threats and errors (variations) and manage human performance. Training based on the results of NOM may, therefore, be a benefit that future research should evaluate.

However, it may be that the ramp environment is not conducive to threat and variation management, simply because ramp crew have comparatively little time (typically less than 45 minutes) or influence to change conditions in their environment. Many of the problems, such as badly presented freight, would be better managed at source via improved practices and procedures in Freight and Baggage services. Alternatively, ramp crew procedures and training for rejecting poorly profiled freight could increase incentives for improving freight and baggage prior to it reaching the ramp.

9.6.5 Limitations of the study

The collection of data was limited in some cases by the operational constraints of the environment. For example, whilst attempts were made to randomise locations and the time of day when observations were conducted, these aspects were heavily influenced by airline schedules, hours of airport operation in each port and the rostering technology. In smaller ports with fewer flights, the observers recorded all the turnarounds that occurred in a week, whereas in larger international ports the observers were restricted to observing turnarounds when local escorts were available, due to local security regulations.

Observations were also limited by the nature of the operating environment itself and what is observable. For example, high levels of ambient noise and the need to wear ear defenders prevented the observers from being able to hear many of the verbal interactions on the ramp. This may be have resulted in lower than expected numbers of communication threats and variations, which were difficult to observe. The communication data, therefore, largely reflects nonverbal forms of communication or communication that occurred further away from the aircraft, such as the read back between the Loading Supervisor and Load Control. It is likely that many communication issues between ramp team members were not recorded.

There may have been some difficulties in interpreting equipment threats separately from driving threats, as some of the ground servicing equipment, or GSE, can also be driven to and from the bay or manoeuvred around the aircraft. Observers may have had difficulty determining whether to code the threats and variations involving motorised GSE as a driving or equipment usage issue. Further clarification of codes for motorised equipment may therefore be necessary.

9.6.6 Possible areas for future study

As with any large dataset, there are many potential avenues for analysis, only some of which could be explored here. For example, the data could be further analysed by port or region, to understand what threats, variations and UOS are more likely in different locations. Further analysis could also consider threat and variations associated with each aircraft type or with different team sizes, for example. Delays could also be analysed in greater depth.

Although analysis of operational delays was not a focus for the current study, there were 23 severe delay events recorded as a UOS. Delay data could be further analysed for associations with other threats and variations. The data could also be compared with other data collected by airlines, such as delay attribution codes. The way in which delays are attributed may be another important factor not studied in this research, and may influence decision making such whether to delay flights by offloading problematic freight (as this may result in a delay penalty being attributed to their team). As delay minutes are particularly costly to airlines, this information could be further investigated to determine aspects affecting operational efficiency. The use of NOM data to provide operational efficiency data could further increase its value and acceptability to the ground handling industry.

9.7 Conclusion

Chapter 9 has attempted to provide an overview of the types of quantitative and qualitative data collected and offer some high-level analysis of the data NOM can contribute to understand threats, variations from SOPs and safety risks in ground operations. There are, of course, many possible interpretations of the data that were not explored in the current study. Nevertheless, we can conclude that NOM was able to provide novel, relevant data about the nature and frequency of threats, variations and undesired states on the airport ramp, as well as some preliminary hypotheses as to why they may be occurring and potential targets for improvement.

NOM data fulfils many of the objectives identified throughout this research for an ideal data collection tool for ground safety. Data from NOM, such as that presented in this chapter, could provide a leading indicator of both human and safety performance for the ground handling industry, and possibly beyond. Observations consider the human and system factors, as well as how operators normally manage risks in every day operations. Narratives provide valuable context to help understand why decisions and actions made sense at the time.

NOM can identify successful management behaviours (although rare in this case) as well as behaviours that are more likely to lead to safety concerns. The data provide evidence-based targets for intervention and inform interventions. Ongoing monitoring could be used to indicate whether those interventions have been successful. NOM also provides a great deal of information about where the system could be improved in terms of interdependencies between service providers or departments or ineffective rules and procedures. It also provides targets for training, with examples and case studies from narratives about effective and ineffective behaviours.

The next chapter will explore in more detail the relationships between threats, variations and undesired states and the implications of these findings for the proposed NOM Model. Chapter 11 will provide an analysis of the NOM tool against the evaluation criteria and Chapter 12 will provide a more general discussion of the research overall.

Chapter 10 Relationships between Threats and Variations

Chapter 9 provided an overview of the data collected and general analysis of the threats, variations and UOS and their management rates. Chapter 10 looks in more detail at the relationship between threats, variations and undesired operational states and the implications for the proposed NOM Model. The results provide an example of the kind of analysis possible using NOM data, although this not exhaustive. The chapter first presents analysis to identify which threats and variations and UOSs are statistically likely to occur at the same time; then the nature of these relationships is explored in order to identify high-value targets for intervention. Examples from the narratives are provided for additional context. The implications of the results are discussed and the potential for this type of analysis to improve safety outcomes are explored.

10.1 Identification of Statistically Significant Relationships between Clusters of Threats and Variations

A systematic approach was taken for the analysis of the potentially very large number of relationships between threats and variations that could be explored. To narrow down the approach, the analysis began by looking for significant relationships between major threat categories and major error categories, as shown in Table 24 below. This table shows the percentage of observations in which threats and variation types occurred together, and the odds of each association occurring by chance. The aim was to consider where the odds of any particular threats and variation occurring together are statistically significant. Cross tabulations were analysed, to understand how often each threat and variation combination occurred and the odds ratio for each potential association. This data explored, through odds ratios, the likelihood of any particular combination of threat and error type occurring by chance. Where there was a very low likelihood of the combination occurring by chance (i.e. high odds ratios) this indicated a strong association between a threat and variation subcategory.

Using a series of strategic cross-tabulations, the table provides odds ratio (OR 95% confidence intervals) for each combination of threat and variation category. The shaded blue cells (where odds ratio confidence intervals do not span 1) highlight where association between threats and variations is statistically significant. These significant associations are of interest because they demonstrate where threats and variations cluster together unusually, and where combinations from each category are common.

For completeness, the table also highlights (in pink) the relationships where the particular variations are *less* likely to occur in the presence of the threat type. Whilst it would be interesting to understand why these relationships might occur, they are outside the scope of this thesis and are not explored further here.

It is important to note that the analysis does not necessarily show causation but does demonstrate that some variations are more likely in the presence of certain threats. For each occasion where the presence of a threat category is significantly related to the presence of a variation category in Table 24, further analysis was conducted to explore the nature of this relationship.

This chapter describes the analysis conducted and the results. Table 24 shows significant associations were more common for Freight and Baggage threats (associated with six major variation categories) and Communication Threats (associated with five major variation categories). Some variation types are common and occur across many types of threats, e.g. Driving variations and Equipment Usage variations, but some are more likely to occur in the presence of particular threats, such as Bay Management and Operational Pressure threats. The results in the table also showed that there were some areas where no significant relationships occurred, such as manpower and weather threats.

Table 24: Table of Odds Ratios (OR) highlighting significant associations between major threat and variations categories.

Threat Type / Variation Type	Airport Facility Threats:% and (OR)	Airside driving Threats:% and (OR)	Arriving Irregular Load Threats:% and (OR)	Bay Manage-ment Threats:% and (OR)	Communi cation Threats:% and (OR)	Documen tation Threats:% and (OR)	Equip. Threats:% and (OR)	F.O.D. Threats:% and (OR)	Freight/ Baggage Threats:% and (OR)%	Load Control Threats:% and (OR)	Man-power Threats:% and (OR)	Opera-tional pressure Threats:% and (OR)	Weather Threats:% and (OR)
Bay management Variations	14.2% 2.19 (OR) 1.31-3.66	8.4% 1.17(OR) 078-1.76	7.5% 0.96 (OR) 0.34-2.71	7.7% 0.98(OR) 0.23-4.21	10.0% 1.35(OR 0.72-2.54	4.2% 0.50(OR 0.12-2.09	9.6% 1.44, (OR) 0.95-2.2	9.5% 1.37(OR) 0.89-2.09	8.1% 1.05(OR) 0.65-1.67	9% 1.17(OR) 0.52-2.619	10.0% 1.32(OR) 0.39-4.16	24.8% 4.96(OR) 3.05-8.08	5.6% 0.69(OR) 0.91-1.02
Communication Variation	0.7% 0.48(OR) 0.06-3.67	1.0% 0.64(OR) 0.24-1.68	0.%	0.%	3.3% 3.10 (OR) 0.99-9.66	4.2% 3.59(OR) 0.79-16.17	1.6%, 1.34, (OR) 0.5-3.56	1.1% 0.78(OR) 0.25-2.41	1.6% 1.34(OR) 0.47-3.85	0%	0.0%	0%	0.% 0.98(OR) 0.98-0.99
Documentation Variations :% and (OR)	20.9% 0.75(OR) 0.5-1.15	22.2% 0.71(OR) 0.55-0.91	9.4% 0.29(OR) 0.12-0.75	11.5% 0.38 (OR) 0.11-0.127	20.8% 0.76(OR) 0.48-1.19	37.5% 1.82(OR) 0.99-3.29	26.4% 1.09, (OR) 0.84-1.4	25.1% 0.98(OR) 0.74-1.29	33.7% 1.72 (OR) 1.3-2.28	32.1% 1.42 (OR) 0.86-2.32	23.3% 0.89(OR) 0.38-2.1	16.8% 0.57(OR) 0.34-0.95	38.9% 1.89 0.73-4.93
Door variations	16.9% 2.38(OR) 1.5-3.9	9.9% 1.23(OR) 0.87-1.88	11.3% 1.32(OR) 0.55-3.16	23.1% 3.18(OR) 1.25-8.08	18.3% 2.59(OR) 1.56-4.32	10.4% 1.19(OR 0.47-3.09	11.4%, 1.57(OR) 1.06-2.3	13.1% 1.92 (OR) 1.3-2.84	11.7% 1.50 (OR) 0.99-2.28	6.4% 0.68 (OR) 0.27-1.73	10% 1.14(OR) 0.34-3.81	9.7% 1.11(OR) 0.58-2.14	11.1% 1.28(OR) 0.29-1.03
Driving Variations	74.3% 0.92(OR) 0.62-1.37	83.4% 2.45(OR) 1.88-3.18	79.2% 1.24(OR) 0.63-2.44	69.2% 0.72(OR) 0.31-1.67	87.5% 2.40(OR) 1.38-4.19	85.4% 1.92(OR) 0.85-4.33	81.2% 1.62(OR) 1.22-2.14	84.7% 2.15(OR) 1.56-2.96	82.5% 1.70(OR) 1.23-2.36	78.2% 1.16 (OR) 0.67-2.02	83.3% 1.63(OR) 0.62-4.28	74.3% 0.93(OR) 0.59-1.44	66.7% 0.64(OR) 0.23-1.72
Equipment Usage Variations	50.7% 0.87(OR) 0.62-1.22	60.4% 1.77(OR) 1.42-2.20	77.4% 3.07(OR) 1.6-5.9	57.7% 1.18 (OR) 0.54-2.59	68.3% 1.98(OR) 1.32-2.95	68.8% 1.94(OR) 1.05-3.61	47.7% 1.28(OR) 1.02-1.61	64.6% 1.87(OR) 1.45-2.39	59.5% 1.37(OR) 1.05-1.77	59.0% 1.25(OR) 0.79-2.0	46.7% 0.75(OR) 0.36-1.55	50.4% 0.87(OR) 0.6-1.28	38.9% 0.54(OR) 0.21-1.43
Loading Variations	14.9% 0.82(OR) 0.51-1.33	15.1% 0.75(OR) 0.56-0.99	1.9% 0.09(OR) 0.01-0.65	11.5% 0.63(OR) 0.19-2.10	20.0% 1.24(OR) 0.78-1.98	25.0% 1.64(OR) 0.84-3.22	19.5% 1.289(OR) 0.95-1.72	17.2% 1.00 (OR) 0.73-1.38	31.1% 3.07(OR) 2.27-4.17	21.8% 1.37(OR) 0.78-2.40	13.3% 0.74(OR) 0.26-2.15	19.5% 1.19(OR) 0.73-1.94	33.3% 2.55(OR) 0.92-6.62

Threat Type / Variation Type	Airport Facility Threats:% and (OR)	Airside driving Threats:% and (OR)	Arriving Irregular Load Threats:% and (OR)	Bay Manage-ment Threats:% and (OR)	Communi cation Threats:% and (OR)	Documen tation Threats:% and (OR)	Equip. Threats:% and (OR)	F.O.D. Threats:% and (OR)	Freight/ Baggage Threats:% and (OR)%	Load Control Threats:% and (OR)	Man-power Threats:% and (OR)	Opera-tional pressure Threats:% and (OR)	Weather Threats:% and (OR)
Read back Variations	14.9% 0.83(OR) 0.51-1.34	15.3% 0.77(OR) 0.58-1.03	0.0% 0.95(OR) 0.94-0.96	3.8% 0.19(OR) 0.03-1.41	33.3% 2.75(OR) 1.82-4.14	31.3% 2.3(OR) 1.23-4.31	17.9% 1.09(OR) 0.81-1.48	17.4% 1.04(OR) 0.76-1.43	30.1% 2.88(OR) 2.13-3.91	15.4% 0.87(OR) 0.46-1.65	16.7% 0.97(OR) 0.37-2.57	8.8% 0.45(OR) 0.23-0.87	22.2% 1.39(OR) 0.45-4.29
Unloading variations	18.2% 1.23(OR) 0.82-2.02	19% 1.91, (OR) 1.39-2.62	43.4% 4.74(OR) 2.69-8.35	26.9% 2.11(OR) 0.87-5.08	23.3% 1.82(OR) 1.16-2.89	2.1% 0.12(OR) 0.02-0.84	15.9% 1.09,(OR) 0.8-1.5	19.9% 1.62(OR) 1.18-2.23	12.0% 0.78 (OR) 0.48-1.04	19.2% 1.36 (OR) 0.76-2.44	23.3% 1.73 (OR) 0.73-4.09	22.1% 1.67 (OR) 1.05-2.69	5.6% 0.33(OR) 0.43-2.45
Restraining variations	13.5% 0.56(OR) 0.35-0.93	18.4% 0.74(OR) 0.57-0.97	0.0%	3.8% 0.15(OR) 0.02-1.11	20.8% 1.01(OR) 0.63- 1.59	45.8% 3.43(OR) 1.91-6.16	21.9% 1.12(OR) 0.84-1.48	23.4% 1.25(OR) 0.93-1.67	33.0% 2.42(OR) 1.81-3.23	24.4% 1.24(OR) 0.73-2.13	10% 0.42(OR) 0.13-1.39	16.8% 0.75(OR) 0.45-1.26	38.9% 2.47(OR) 0.95-6.43

Some major threat categories appeared to be associated with lower incidents of major variation categories, indicated by the lighter shaded pink cells in Table 24. This indicates that the presence of airside driving, arriving irregular load, bay management and operational pressure threats, was associated with a lower incidence of documentation variations. Similarly, airside driving and arriving irregular load threats were associated with a lower incidence of loading variations. Operational pressure threats were associated with lower read back variations, and airside driving threats were associated with lower incidence of restraining variations.

The nature of these apparently protective associations may be features of the methodology. For example, documentation threats were associated with lower unloading variations. In this case, if the documentation threat was the lack of an Off-Load Instruction Report (OLIR), it would not be possible for the observer to record any unloading discrepancy variations without reference to the documentation. However, for some protective associations, the nature of the relationship is not known. As this was not the subject of the current study they are not considered further here, but maybe of interest for future research as discussed in chapter 12.

10.2 Analysing Subcategories within Significant Clusters

The next stage of the analysis was to investigate which particular subcategories of threats and variations contributed to each significant relationship. For each of the significant relationships shown in blue in Table 24, a series of tables were made comparing individual threat and variation subcategories.

This information made it possible to create a series of diagrams to depict threats and variations that were strongly associated with each other. For each of the diagrams, threats are shown in yellow and variations in blue. The thickness of the connecting lines between the boxes indicates the strength of the relationships in terms of the increased odds that this particular threat and variation are related.

As well as investigating the relationships between threats and variations, it is also possible to use the data to investigate relationships between variations and undesired states and attempt to understand why they might occur. A similar process was followed to identify variations and threats that occurred at the same time as UOSs were present. However, due to the relatively low number of UOSs, it was not possible to calculate reliable odds ratios. Nevertheless, the association of undesired states with their preceding variations and threats is also represented diagrammatically. In these diagrams UOSs are shown in orange, and the connecting lines indicate the number of times the UOS was preceded by each variation or threat type. The next section explores the some of the associations discovered and displayed in the diagrams.

10.2.1 Airport facility threats and variations

There were 169 airport facilities threats, affecting 11.4% of observations. Analysis of the specific subcategories of both airport facility threats and each of these variation types showed that there were three airport facility subcategories that were significantly associated with two variation subcategories, one relating to bay management variations and the other to door variations. Addressing these particular airport facility threats should, therefore, reduce the odds of these variations occurring.

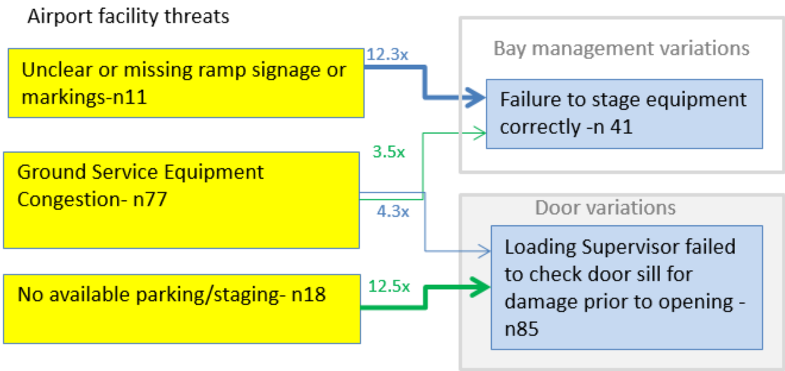


Figure 54:Airport Facility threats and their associated variations.

Figure 545 shows an association between the threat of unclear or missing ramp signage and failure to stage equipment correctly. Failure to stage equipment correctly was 12.3 times more likely to occur when ramp signage or markings were not in place. Congestion was a fairly common form of airport facility threat (n77). The presence of congestion was associated with a 3.5 times increase in the failure to stage equipment correctly, and a 4.3-fold increase in the loading supervisor failing to check the door sill for damage. Lack of parking or staging space was observed as a threat on 18 occasions. This was associated with a 12.5-fold increase in the variation of failing to perform a door sill check. The relationships demonstrated in Figure 545 indicate that congestion on the ramp in general may have previously unrealised ramifications for the conduct of other operational tasks. Addressing the congestion issue would therefore be an important area for safety intervention.

Examples from narratives:

> *'Many of the passengers were then walking in unmarked areas and there were no other passenger controllers positioned between the aircraft and the terminal'.*
>
> *'There are no visible equipment staging markings on this bay, although there are equipment storage lines, these however equipment had parked in them creating congestion'.*
>
> *'The equipment clearance lines are very close to the FWD holds of the aircraft. A pallet loader was behind the lines but obstructs [the] tug and catering vehicles when they service the aircraft. The loading supervisor opened the door without checking the sill for damage'.*

10.2.2 Airside Driving threats and associated variations

There were a very high number of airside driving threats overall (n922), which affected 52.2% of observations. Although they were common across all observations, airside driving threats were only significantly associated with particular types of variation in the unloading, equipment usage and driving variations categories.

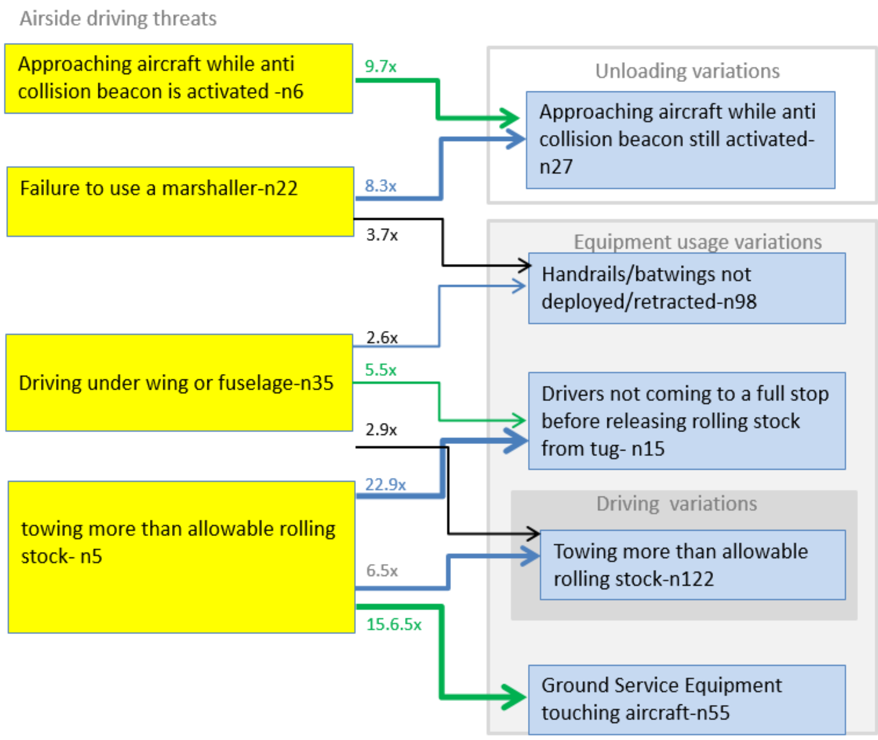


Figure 55: Airside Driving threats and associated variations.

Non-ramp crew towing too much stock was only recorded as a threat on five occasions, but the presence of this threat was associated with a 6.5-fold increase in ramp crew also displaying the same behaviour. This may be because ramp crew take over the driving of already overloaded dollies. This threat increased the likelihood of the variation of failing to come to a full stop before releasing rolling stock by 22 times, and GSE touching the aircraft was 15.5 times more likely to occur when towing excessive loads, possibly due to the effect of heavier weight on braking performance.

Figure 556 shows the threat of non-ramp crew approaching the aircraft before the anti-collision beacon is switched off was associated with the variation of ramp crew also approaching the aircraft too soon. Failure to use a marshaller further exacerbated the ramp crew variation of approaching the aircraft too soon (by 8.3 times) and more than doubled the likelihood that handrails or batwings might be left in the wrong position.

Non-ramp crew driving under the wing or fuselage was a threat observed 35 times during the audit period. This threat was associated with ramp crew operating equipment with batwings or handrails not retracted and increased risk of equipment striking the aircraft.

Examples from narratives provide further context of these issues:

> *'18:18 aircraft arrives and parks on bay the anti-collision beacon is still flashing and the engineer still hasn't given the thumbs when the mobile stairs start to approach the aircraft, the engineer with the guiding paddles motions to the ramp team to stop approaching and she obeys this command'.*
>
> *'A ramp crew positioned the DPL (Disabled Passenger Lift) to the push upstairs. A marshaller was not used'.*
>
> *'The driver arrived on the bay towing a train of six dollies. The belt loader was also driven to the staging area with the side rail in the raised position, the SOP's state that the side rails must be in the lowered position when driven on the ramp. 1316 The aircraft arrived on the bay and was chocked 26 minutes late. The belt loader was in the raised position prior to the loader being positioned at the aircraft, the SOP's state that the belt is not to be raised until making final adjustments to the aircraft'.*
>
> *'The belt loader used to unload the bulk hold has both safety rails up and one side bat wing deployed, the other side bat wing was not fully deployed… The belt loader is removed from the aircraft, the hand rail is still in the raised position while it is being driven away'.*
>
> *'The driver has taken the barrow of cargo to the bulk hold and dropped it at [the] end of the belt [loader]; the driver has then driven under the fuselage as a short cut to connect to two low Profiles with cargo and taken them to the forward hold'.*

10.2.3 Arriving irregular load threats and associated Equipment variations

There were 64 arriving irregular load threats, affecting only 4.1% of observations. When they did occur, arriving irregular load threats were significantly associated with equipment usage and unloading variations. If arriving irregular loads could be reduced at the port of origin, this may lead to a decrease in some of the variations shown in Figure 567.

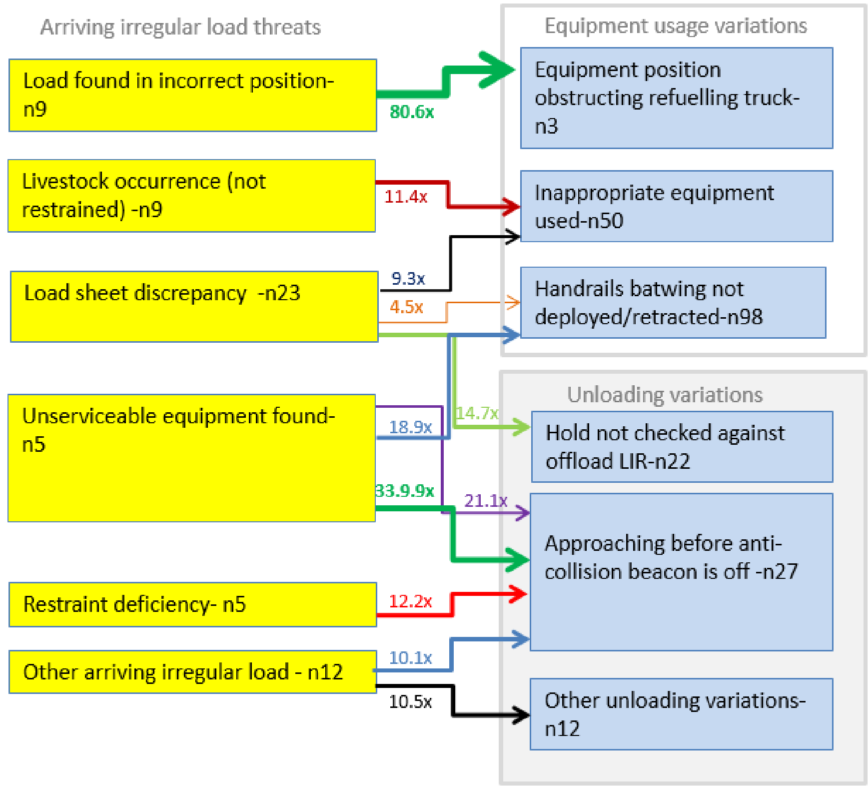


Figure 56: Arriving Irregular Load threats preceding Equipment and Unloading variations.

Irregular loads are encountered by the ramp crew when they come to unloading the aircraft. Finding an irregular load was associated with an increase in equipment usage variations and unloading variations by the ramp crew. Finding the load in an incorrect position was only recorded in nine observations, but when it occurred it led to an 80-fold increase in the variation of equipment obstructing the refuelling truck. Inappropriate use of equipment increased by 11.4 times when there was also a livestock occurrence threat present, such as an unrestrained animal in the hold. Arriving irregular loads threats were associated with a 10-fold increase in the unloading variations of not checking the hold against the Off-Load Instruction Report and approaching the aircraft while the beacon is still on (n27).

Figure 54 also shows that when there was a load sheet discrepancy from Load Control (such as information missing or OLIR not provided), the failure to check the hold against the Offload Instruction Report was 14.7 times more likely. When the threat of unserviceable equipment was present, ramp crew were 18 times more likely to use inappropriate equipment instead. The use of incorrect equipment also increased by 4.5 times when the threat of a discrepancy with the arriving load (n23) was present.

Examples from narratives provide further context of these issues:

> *'Three AVI are found unsecured upon opening the door. There is no tie down restraints supplied by cargo, and the animals are completely unrestrained'.*
>
> *'The irregular load is discovered during the unload 2 bags of used headsets are removed from Hold 11, these are not shown, on the inbound LIR, therefore leading to an Irregular Load'.*

10.2.4 Bay Management threats and associated variations

There were only 27 bay management threats, affecting 2% of observations; however, bay management threats were significantly associated with door variations.

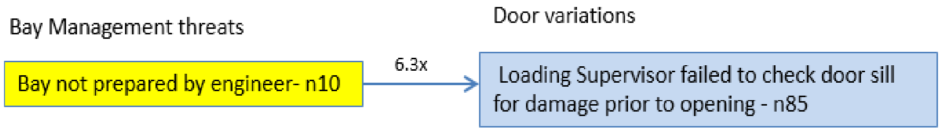


Figure 57: Bay Management threats and door variations.

In Figure 578, for example, the bay not being prepared by the engineer was associated with a 6-fold increase in ramp loading supervisors failing to ensure that the door sill and surrounds were free of damage prior to opening (n85). Examples from narratives provide further context of this issue:

'The LAMEs had not prepared the bay by 10:49 the aircraft is on chocks at 10:55, ten minutes early'.

'The rear hold was unloaded without the loading supervisor performing a check on the sill for damage. The ramp team driver and an in-hold operator open both cargo doors, neither check the door or sill for any damage as is required'.

10.2.5 Communication threats associated with variations

There were 124 communication threats, affecting 9.2% of observations. Communication threats were associated with four different major categories of variations including door, driving equipment usage, read back and unloading variations. Reducing communication threats could, therefore, reduce the likelihood of a wide range of variations in these areas.

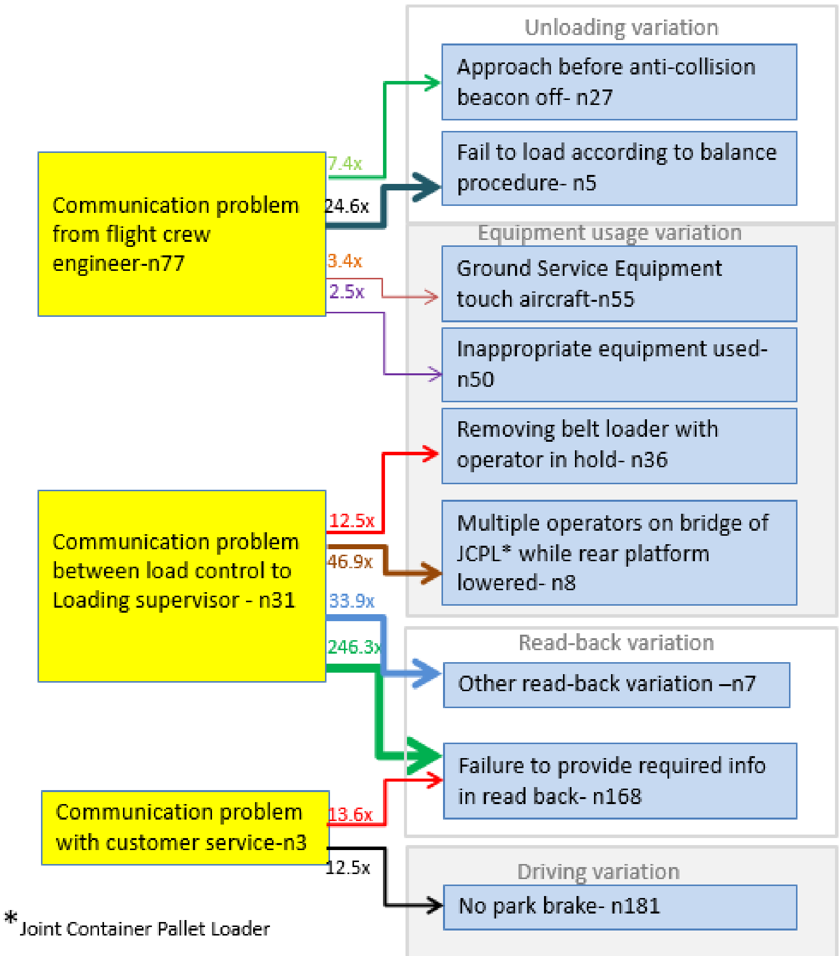


Figure 58: Communication threats and associations with variations.

In many cases, communication threats greatly increased the odds of particular variations. For example, communication problems between load controls to the Loading Supervisor led to a 246.3-fold increase in ‘other’ read back variations, as the read back usually repeats the information provided by the Load Control. The same threat was also associated with a 33.9-fold increase in the number of read back variations generally. Failure to provide the required information in the read back was also 13.6 times more likely when there was a communication threat present with customer service personnel, and the same threat also appeared to increase by 12.5 times the likelihood of parking brake variations by ramp crew.

This may in some cases be due to problems with communication equipment. Communication threats occurring between the flight crew and the engineer were recorded 77 times. Typically, the engineer provides the ‘thumbs up’ signal to communicate when it is safe to approach the aircraft. When this threat was present it was associated with 7.4-fold increase in approaching the aircraft before the beacon is off. The same threat was also associated with a 24.6-fold increase in the failure to load according to the balance procedure, a 3.4-fold increase in GSE touching the aircraft, and more than double the variations of ramp crew using inappropriate equipment.

Examples from narratives provide further context of these issues:

> *‘There was no signal to approach the aircraft given by the engineers’.*
>
> *‘At 1314 normal ACARS reporting system does not work here so the final load and trim figures are given directly to the captain after the loading supervisor obtains baggage and passengers position figures’.*
>
> *‘Customer services failed to contact the Loading Supervisor and notify him about how many pieces are being sent down the chute or what the object was. The drum has been loaded on the baggage chute without communication… the drum rolled down the chute and collided with the push out vehicle’.*

10.2.6 Documentation threats and associated variations

There were 49 documentation threats, affecting 3.7% of observations. Documentation threats were particularly associated with two types of variation: read back variations and restraining variations. Addressing incomplete or incomplete documentation threats should decrease the odds of these variations occurring.

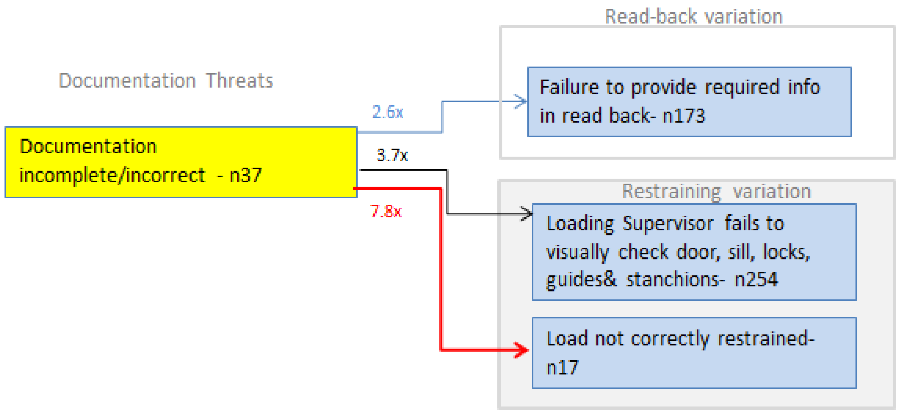


Figure 59: Documentation threats and associated variations.

The threat of incomplete or incorrect documentation occurred 37 times during the observation period. When it occurred, its presence was associated with a 2.6-fold increased likelihood of the variation of failing to provide the required information during the read back. This is most likely because the Loading Supervisor does not have a copy of the correct information to read back. The same threat, however, was also associated with two restraining variations, which may result in the load shifting in flight. Incomplete documentation was associated with a 7.8-fold increase in incorrectly restrained loads (n17). Documentation threats also increased by 3.7 times the likelihood of the Loading Supervisor failing to check the locks guides and stanchions correctly (n544).

Examples from narratives provide further context of these issues:

> *'The loading supervisor was present but none of the ramp team had a copy of the Load Instruction Report'.*

'I observed an AVI being loaded into position 21 (instead of 11) as per the LIR with baggage surrounding the crate it was not secured with approved straps'.

'The belt loader operator secures the webbing and stanchions at the forward hold then closes the door. The Loading Supervisor was not present for this and did not check the forward Hold as required by the Aircraft Loading Checklist'.

10.2.7 Equipment threats and associated variations

Equipment threats were common (n 534), affecting 34.3 % of observations. These threats were particularly associated with driving variations and equipment usage variations. Reducing equipment threats should decrease the risk of these associated variations.

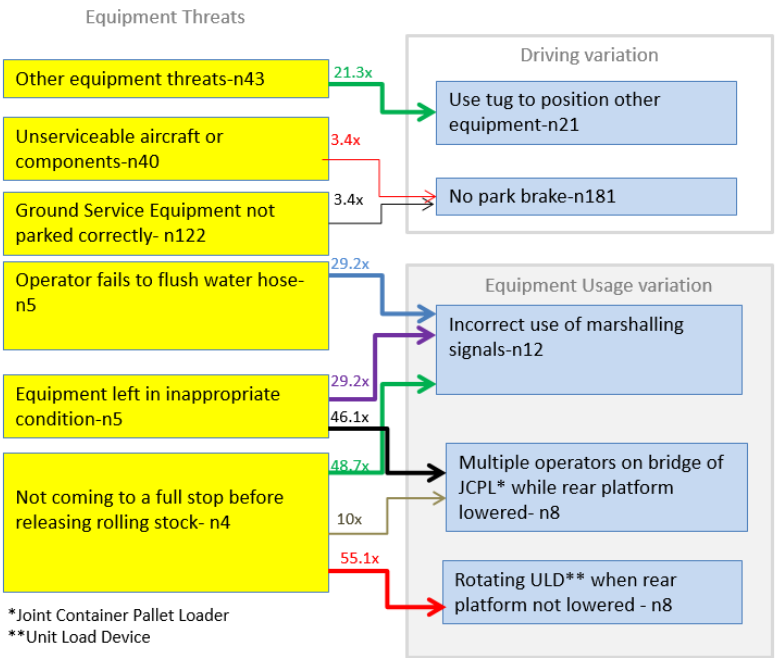


Figure 60: Equipment threats and associated variations.

Equipment threats increased the likelihood of equipment variations and driving variations occurring as shown in Figure **60**61. Non-ramp crew failing to come to a full stop before releasing rolling stock (n4) was associated with a 55.1-fold increase in variations where multiple operators were seen on the bridge of a JCPL while the rear platform was loaded. This was also found to increase by 46 times when ramp crew found equipment in an inappropriate condition (n5).

Unserviceable aircraft components were associated with a 21.3-fold increase in the driving variations and an increase in failing to use a parking brake by 3.4 times. The threat of non-ramp crew failing to flush the water hose was associated with a 29-fold increase in the variation of using incorrect marshalling signals. 'Other' Equipment threats led to an associated increase in the dangerous use of tugs to position equipment. The use of the 'other' category suggests that the codes were not specific enough to identify the threats related to jammed equipment or vehicles that may be associated with this practice. Equipment threats were associated with a 21.3 times increase in the variation of using a tug to position other equipment.

Examples from narratives provide further context of these issues:

> *'The rear tug driver was using his tug's front bumper to push a low profile forward as he wanted to get the dollies that were parked behind the low profile. The tug sustained minimal damage'.*
>
> *'A dolly cannot be rotated so the loading supervisor calls for the tug driver to push the tug against the dolly to force it round'.*
>
> *'The belt loader at the front [of the aircraft] has defective brakes, as the staff was placing a chock behind the rear wheel so as to stop the belt rolling back. [However] they did not place a chock in front of the wheel so the belt loader was able to move forward and damage the aircraft'.*
>
> *'A driver from the baggage room delivers a barrow with 1 piece, he fails to come to a complete stop before releasing the barrow, and it travels for approx. 2m before it stops'.*

'The AFT JCPL operator had come up on the bridge of JCPL to assist forward hold operators to manage the equipment failure'.

10.2.8 Freight and baggage threats and associated variations

There were 385 freight and baggage threats, affecting 23.7% of observations. Freight and baggage threats were associated with six different major variation types, including documentation, driving, equipment usage, loading, read back and restraining variations. Reducing the occurrence of freight and baggage threats before they arrive at the aircraft bay should also decrease the odds of a wide range of variations occurring during the turnaround. As expected these threats were particularly associated with loading variations, as demonstrated in Figure **61**61, and documentation variations, as shown in Figure 622.

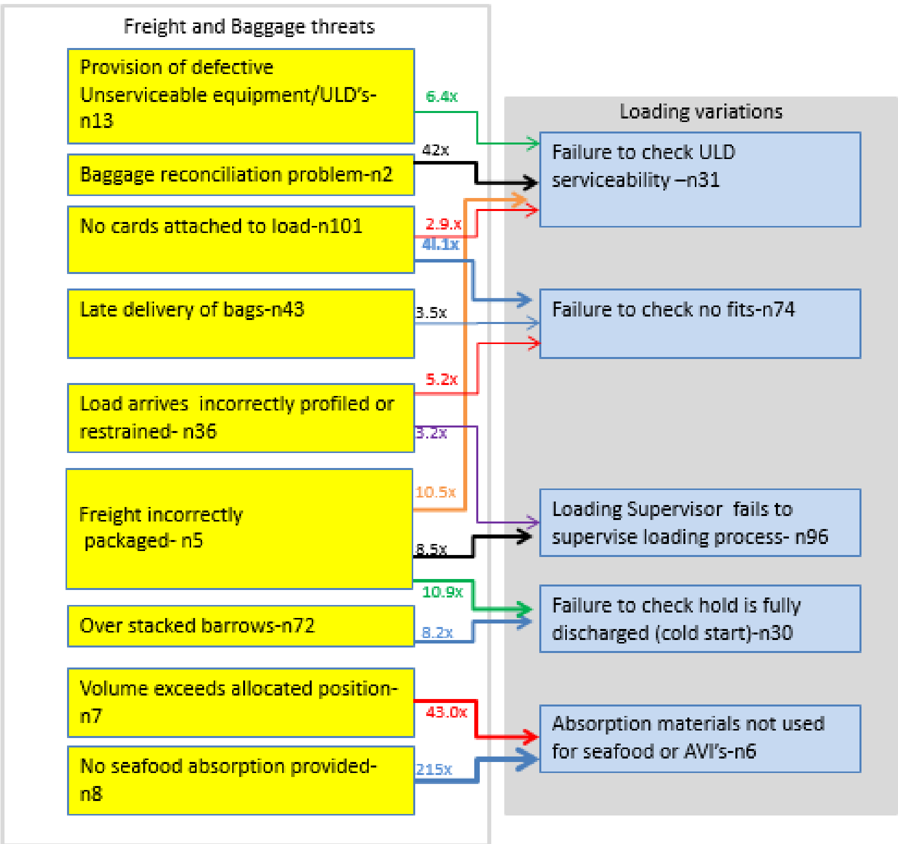


Figure 61: Freight and baggage threats and associated loading variations.

When unserviceable ULDs are sent from Freight and Baggage, this was also associated with a 6.4 times increase in the variation of failing to check for ULD serviceability, reducing the likelihood that the ULD will be removed before it travels. Having no card attached to the load also doubled the likelihood of failing to check ULD serviceability. Other threats also appeared to be associated with the variation of failing to check ULD serviceability. For example, the threat of a baggage reconciliation problem was associated with a 42-fold increase in this variation.

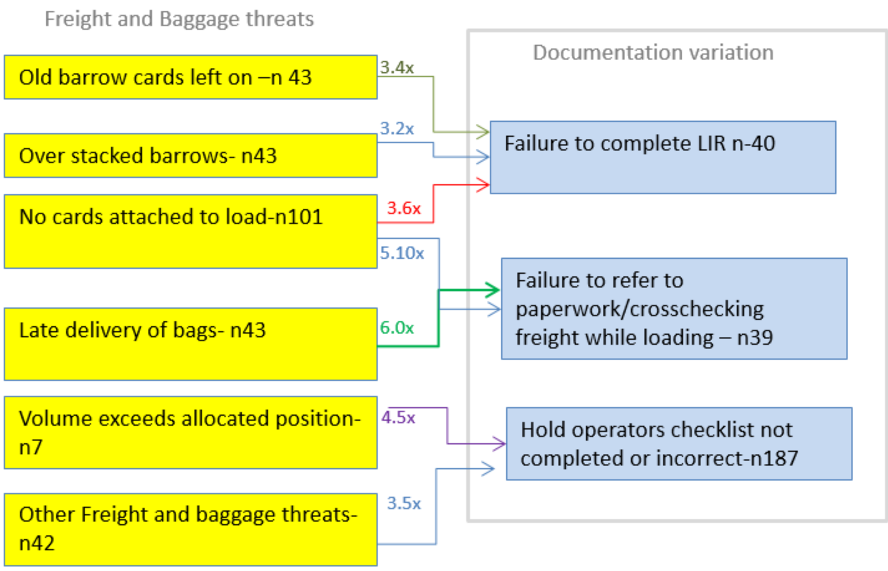


Figure 62: Freight and baggage threats and associated documentation variations.

Freight and baggage threats also increased the likelihood of documentation variations, equipment usage variations and loading variations occurring. For example, the threat of old barrow cards being left on freight was associated with a 3.4 times increase in incomplete Load Instruction Reports (LIR). The threat of having no cards attached to the load was also associated with a 5-fold increase in paperwork variations and a 3.2-fold increase in incomplete LIR.

The late delivery of bags was associated with a 3-fold increase in failing to check the hold for 'no-fits' (empty positions). The load arriving incorrectly profiled was associated with a 5.2-fold increase in failing to check for no fits, and a 3.2-fold increase in failure to supervise the loading process. The threat of over-stacking the barrows was also associated with a 3.6-fold increase in not completing the LIR. Failure to check that the hold is fully discharged was 10 times more likely when the freight was incorrectly packed in the first place, and eight times more likely when it was over-stacked.

When load volumes exceeded their allocated position, it was associated with a 43-fold increase in failure to use absorption materials for the transport of seafood or animals, perhaps because there was not enough space. Of course, if the absorption materials are not provided it is extremely likely that the freight will be transported without the absorption material in place, as indicated by the 215-fold increase shown in Figure 62.

Examples from narratives provide further context of these issues:

> *'The ULD was observed with a bulge in the side. The Loading Supervisor was observed making a call but the conversation was not heard. The bulging ULD was then loaded with the bulge against the side of the aircraft'.*
>
> *'The Loading Supervisor inspected the freight and was observed cross-checking the freight against the LIR. He collected and removed all the barrow cards from the barrows but did not identify the box which was on the incorrect barrow. The box was eventually loaded into position 31 of the aircraft when it should have been loaded into position 21'.*
>
> *'Loading starts at 15.09 as the load is being towed out there are bags piled on the barrows higher than the end rail. Three pieces of baggage then fall from the barrow'.*
>
> *'Again the Loading Supervisor leaves the bay while the rear is still being loaded; he has gone back to the allocator's office to call through the load. Just prior to the call back, the tug driver advises the Loading Supervisor that another bag was placed in the aircraft'.*

10.2.9 Operational pressure threats and associated variations

There were 116 operational pressure threats, affecting 8.7% of observations. Operational pressure threats were particularly associated with bay management variations. These are interesting in that they suggest reasons for the pressure. For example, early aircraft arrivals can be problematic if the crew are not yet in place, whereas late arrivals compress turnaround times for loading and unloading tasks. Late departures can also present flow-on operational threats if crews are delayed getting to the next scheduled turnaround.

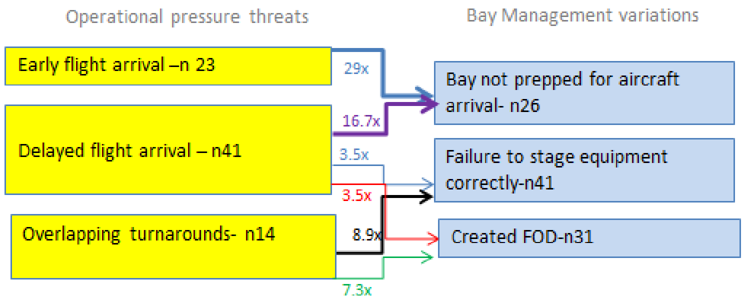


Figure 63: Operational threats associated with bay management variations.

Figure 634 shows that early flight arrival threats were associated with a 29-fold increase in the likelihood of the bay not being prepared for the aircraft's arrival. However, a delayed flight arrival threat was also associated with a 16.7 times increase in the bay not being prepared, and a 3.5-fold increase in the failure to stage equipment correctly. The threat of overlapping turnarounds was associated with an 8.9-fold increase in failure to stage equipment correctly, and a 7.3-fold increase in the creation of Foreign Object Debris (FOD) on the bay.

Examples from narratives provide further context of these issues:

> *'After arriving on bay, the PDA informs the Loading supervisor they are now to start loading another aircraft and the whole ramp team leave to locate the LIR'.*
>
> *'The loading supervisor arrived on the bay but no other ramp members were available as they were unloading another aircraft at the international terminal'.*

10.3 Variations Associated with Undesired Operational States

UOSs are comparatively rare in the data set, as would be anticipated. There were only 165 undesired states recorded overall, 102 of which were associated with at least one variation. Using NOM data, it is possible to identify which variations are associated with these undesired operational states; however, as the numbers are small, no odds ratios are shown.

10.3.1 Airside Driving variations and associated UOSs

Airside driving variations were common (n2290), affecting 75.7% of all observations. Airside driving variations preceded undesired states of conflict or near collision between aircraft and ground servicing equipment (GSE), or between two ground servicing vehicles. Reducing airside driving variations such as speeding or driving underneath the wing, for example, could reduce the opportunity for these events to occur.

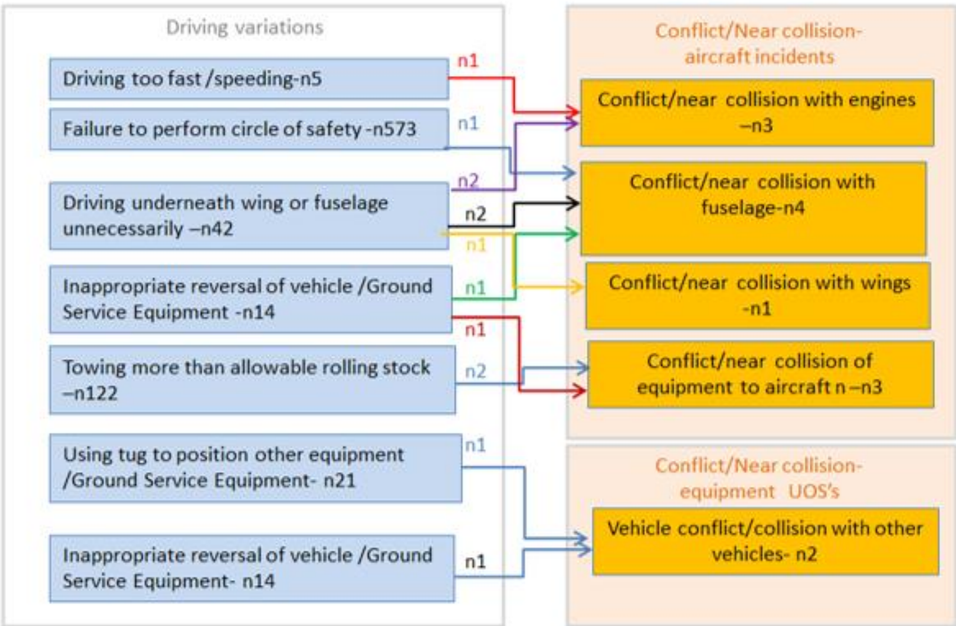


Figure 64: Driving variations associated with UOSs.

There were three UOS events of conflict or near collision with engines. One of these was preceded by a speeding variation, and two were preceded by driving underneath the wing or fuselage unnecessarily. Conflict or near collision with the fuselage was recorded four times. One of these was preceded by a failure to perform the circle of safety procedure, two were preceded by driving under the wing or fuselage unnecessarily, and on one occasion this was preceded by the inappropriate reversal of a vehicle. Conflict or near collision with equipment to an aircraft (n3) was proceeded by one inappropriate reversing variation and two variations where crew were towing more than the allowable rolling stock. The UOS of two vehicles being in conflict or colliding with each other was recorded on two occasions. One was preceded by the use of tugs to reposition equipment, and one was associated with an inappropriate reversal of a vehicle.

Examples from narratives provide further context of these issues:

> *'The mail driver then picked up three low profiles for the offload of the ULD's at the FWD hold; he drove directly under the fuselage to align with side of the FWD JCPL. There was very minimal room for the driver to safely place these low profiles as the catering truck was in place, so a second tow motor driver came into the picture. As the driver slowly drove between the side of the JCPL and the catering truck the second driver pushed each low profile as the first driver drove through. At one point, while reversing, the [second] driver caused the rear of the tow motor to come in contact with the aircraft's nose wheel, this was done at a slow speed And caused nil damage to the wheel or tyre of the aircraft...'.*
>
> *'The second trailer was loaded first and the Loading Supervisor then reversed the trailers to load the first one. When doing this, the first trailer made contact with the JCPL extension deck. The trailer was towed forward and reversed again'.*

10.3.2 Variations associated with Loading UOSs

A range of variations preceded loading UOSs, but the most common category was loading variations. Prevention of loading events should prioritise reducing the variations shown in Figure 656 (over). There were 14 UOS incidents where unserviceable ULD were loaded onto the aircraft; of these, two were preceded by use of the defective ULD in the first place and 12 by occasions where the ramp crew failed to check for ULD serviceability. There were nine occasions on which the load was transported in the incorrect position. On six occasions this was preceded by ramp crew loading the cargo into the wrong position in the first place, and one example of the Loading Supervisor failing to supervise the loading processes until the load was on board. The same failure to supervise variation also preceded an irregular load on two other occasions.

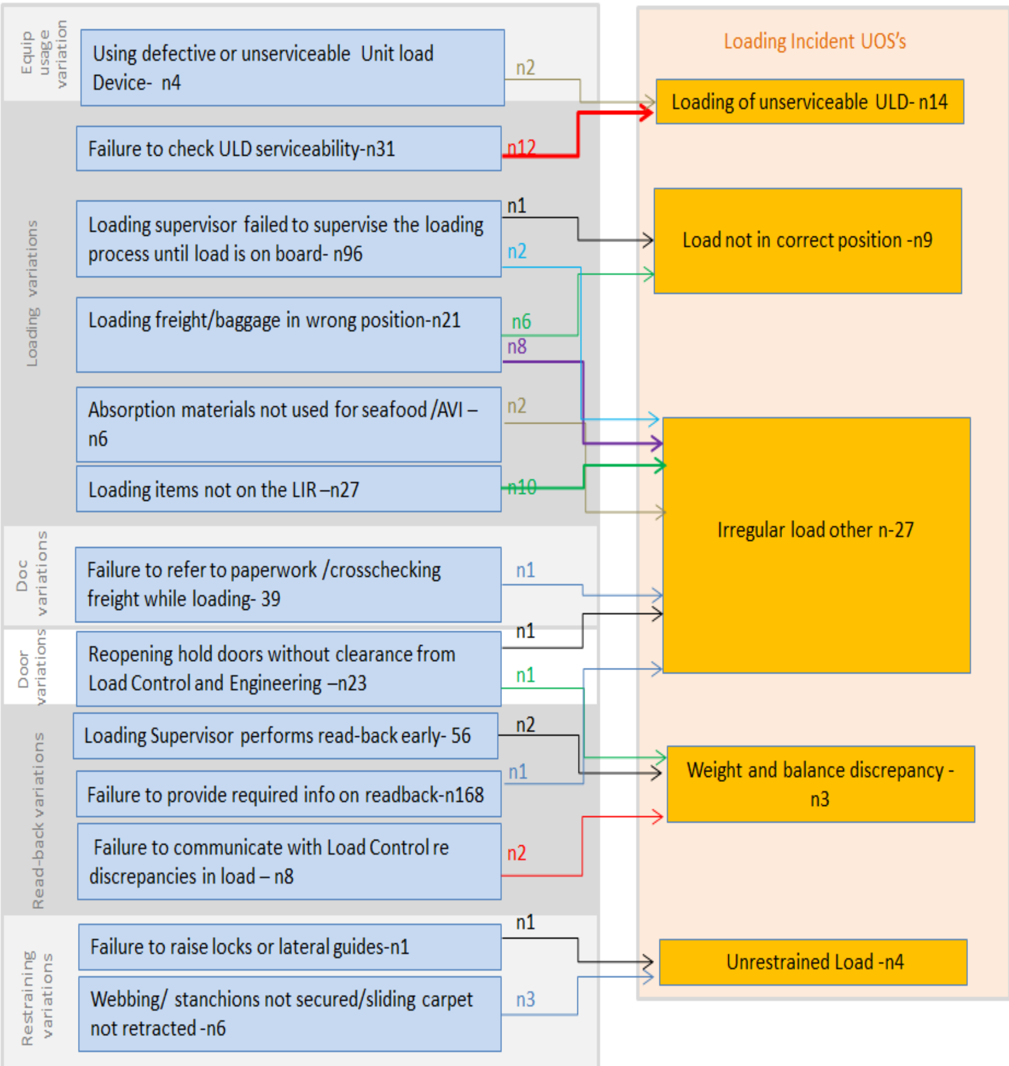


Figure 65: Variations preceding Loading undesired states.

Loading items that were not listed on the Load Instruction Report was associated with a 10-fold increase in irregular loads. Failure to refer to the paperwork while cross-checking freight also preceded an irregular load on one occasion, as did reopening the doors without clearance from Load Control.

There were three weight and balance events, two of which were preceded by read back variations where the crew did not challenge Load Control regarding the discrepancy. There were four unrestrained loads, on three occasions the proper restraints were not secured or retracted, and on one occasion there was a failure to raise the locks and lateral guides. Examples from narratives provide further context for these issues:

> *'At the rear of the aircraft a container was on the bay waiting to be loaded with a damaged top curtain rail. This was quite visible but still loaded onto the aircraft'.*
>
> *'I arrived on the bay at 1840, I have a copy of the LIR for this flight, and it shows AVIs in position 21. [The] AVIs are to be planned for position 11, as per the weight and balance standing order to load control'.*
>
> *'The forward hold is being loaded first, starting with the freight followed by the first 70 bags, the tug driver has already bought out these bags and counted them to ensure that there are in fact 70 bags on the first 2 barrows, therefore the remaining bags can be taken to the rear hold for loading. He advises his Loading Supervisor of this, however by doing this there was some priority bags loaded into position 21 that should have been loaded last in position 31'.*
>
> *'A late bag is carried out by a ramp member and placed in the forward hold; no prior permission was given to reopen the hold door by the engineer, as he was watching the rear hold door being closed. Late bag wasn't included in the read back'.*

10.3.3 Variations associated with livestock UOSs

Equipment usage variations and restraining variations preceded several undesired states concerning the transportation of livestock. Addressing these variations may improve the likelihood that animals will be safely restrained and transported.

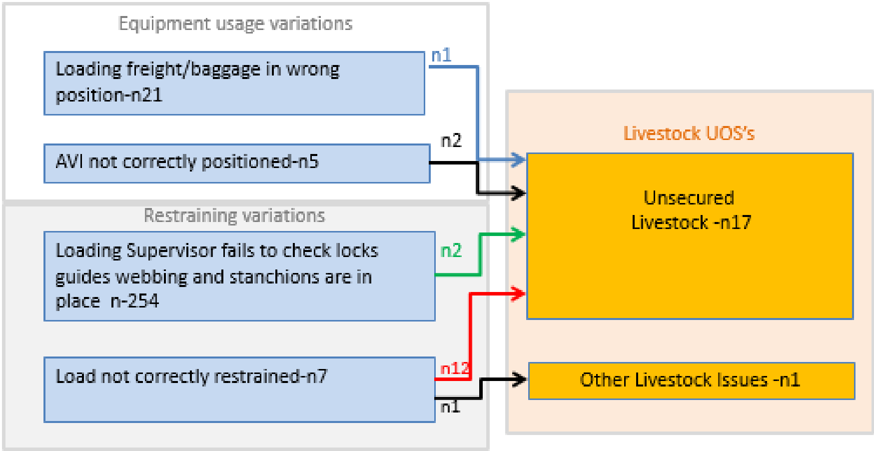


Figure 667: Variations preceding livestock UOSs.

Figure 667: 7 shows that of the 18 livestock incidents, 17 were unsecured livestock with one 'other' livestock issue. Unsecured livestock issues were preceded by 12 variations where the load was not correctly restrained, two occasions on which it was incorrectly positioned and two occasions on which the Loading Supervisor failed to check the locks, guides, webbings and stanchions were in place before take-off.

> *'At 18.03 the forward hold has finished loading and the in hold operator has removed the belt loader and parked it in its designated parking area. On further investigation the two AVI livestock that had been loaded into position 11 had not been restrained'.*

10.3.4. Variations associated with aircraft conflict UOSs

Preventing equipment usage and unloading variations that allow equipment to come into contact with aircraft doors and sills may reduce the potential for aircraft damage.

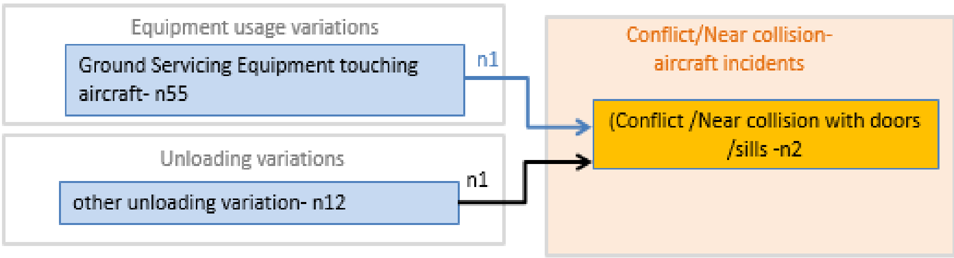


Figure 67: Variations preceding conflict/near collision of equipment with aircraft UOSs.

Figure 678 indicates that collision with doors and sills was preceded by one occasion where the GSE was touching the aircraft and one unloading variation. Examples from narratives provide further context for these issues:

> *'The mobile stairs were marshalled into final position by the Movement Controller (MOCO) and the rubber bumper on the stairs was touching the aircraft. There was no required gap of 50mm between the bumper and the aircraft'.*
>
> *'The first container in the doorway of the forward hold is removed, the operator is not watching the container, as it comes out of the doorway the ULD spins to the right and impacts with the door sill, the aluminium are flattened and a proper seal with the door rubber will now not be possible. The operator does not inform anyone of this and continues unloading'.*

10.3.5 Variations associated with ground event UOSs

Driving variations were common across all observations (n2290); however, two specific driving variations were found to precede the four ground event UOSs.

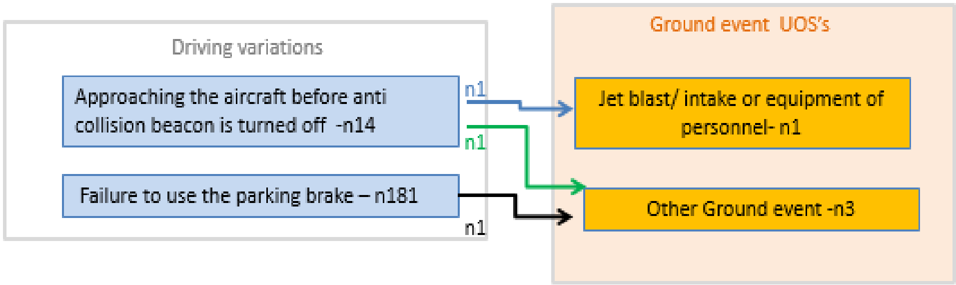


Figure 68: Variations preceding ground event UOSs.

Figure 68 shows that of the four ground incidents observed, three were recorded as 'other' ground events and one was recorded as a jet blast incident. The jet blast event was preceded by the variation of approaching the aircraft before the anti-collision beacon was turned off. The three other ground events were proceeded by the variation of approaching the aircraft too soon, but also by one example of a failure to use a park brake and another bay management variation.

> *'An A380 on the adjoining bay was in the final stages of departure… as the aircraft taxied passed the bay and turned towards the runway, the ramp team on the bay were hit with jet blast'.*

10.3.6. Variations associated with dangerous goods UOSs

The incorrect restraining of the load preceded two UOS events of unrestrained dangerous goods. Addressing restraining variations may reduce the opportunity for this UOS to occur.

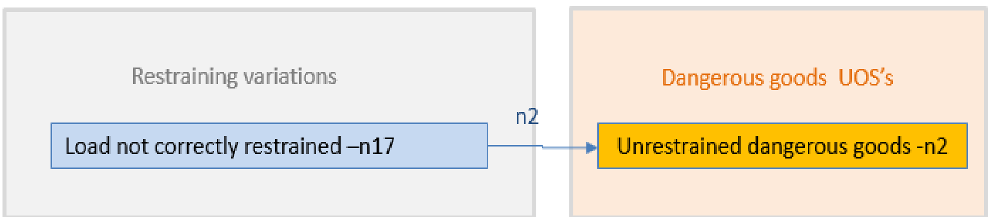


Figure 69: Restraining variations preceding dangerous goods UOSs.

There were two dangerous good incidents recorded, as shown in Figure 69; one involved the loading of dangerous goods only permissible on cargo aircraft, and the other incident involved unrestrained dangerous goods. There were 17 variations of unrestrained cargo, two of which directly preceded an unrestrained dangerous goods incident. An example from narrative provides further context:

> *'The dry ice was identified, but not as a "Dangerous Goods" on the Load Instruction Report. Load Control should not have planned the dry ice to be loaded in the FWD hold with AVIs'.*

10.3.7 Threats associated with delay UOSs

There were 41 threats associated with undesired states where no variation was recorded. Half of these were associated with delay events, most of which are imposed by other departments or service providers and impact the ramp crew. If addressing the causes of significant operational delays is a priority, then addressing these threats should reduce the opportunities for delay events.

Figure 70 (over) shows there were six occasions where operational pressure threats were associated with severe delays, five threats originating from the Freight and Baggage areas and two originating from the Load Control. Addressing the specific interface issues that contribute to delays could therefore improve on time performance.

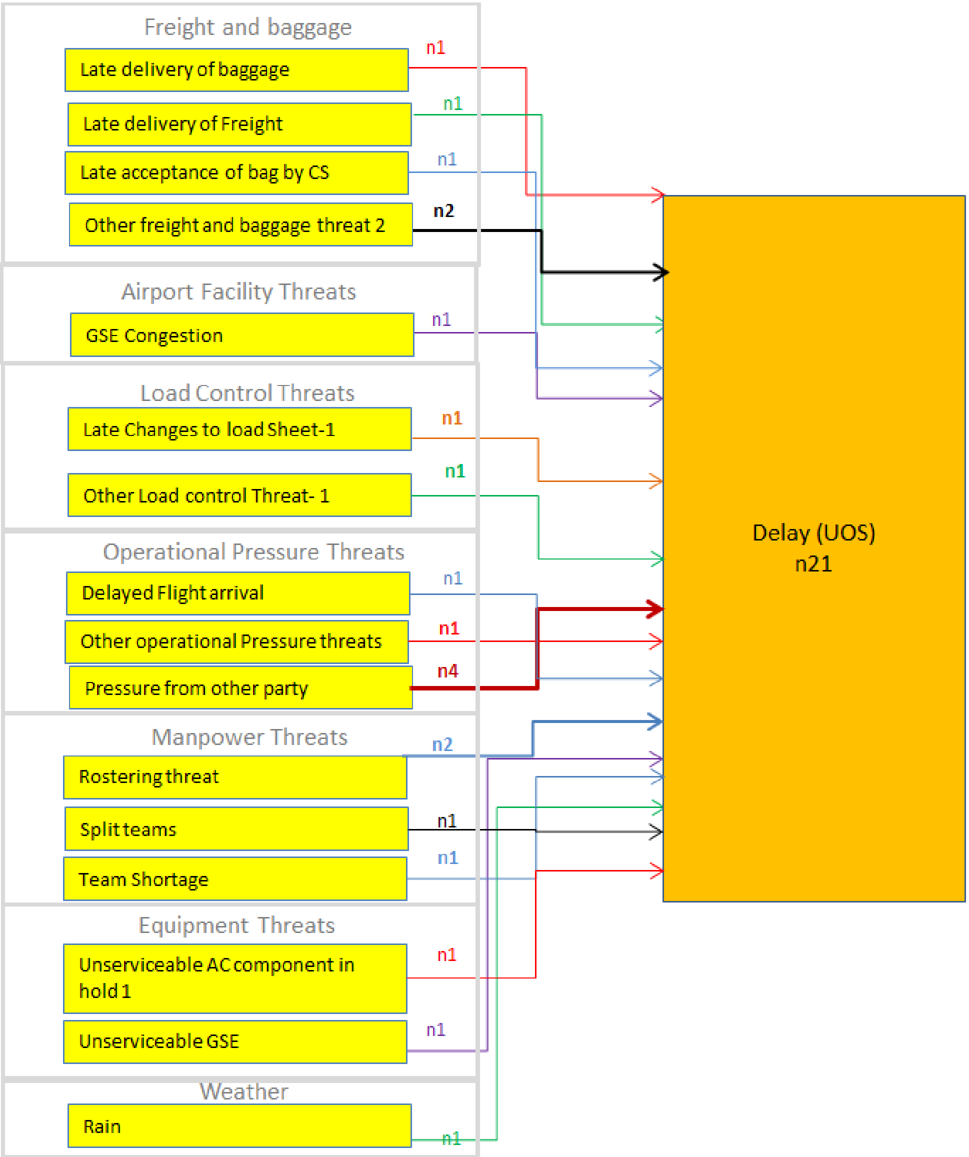


Figure 70: Threats associated with delay UOSs.

10.3.8 Threats associated with conflict/near collision with aircraft

There were six UOS events involving a conflict or near collision with the aircraft. These were associated with threats from non-ramp crew's use of ground servicing equipment near the aircraft. Given the industry's need to reduce incidents of aircraft damage, as discussed in Chapter 2, addressing all occasions where GSE may come into contact with the aircraft should be a priority.

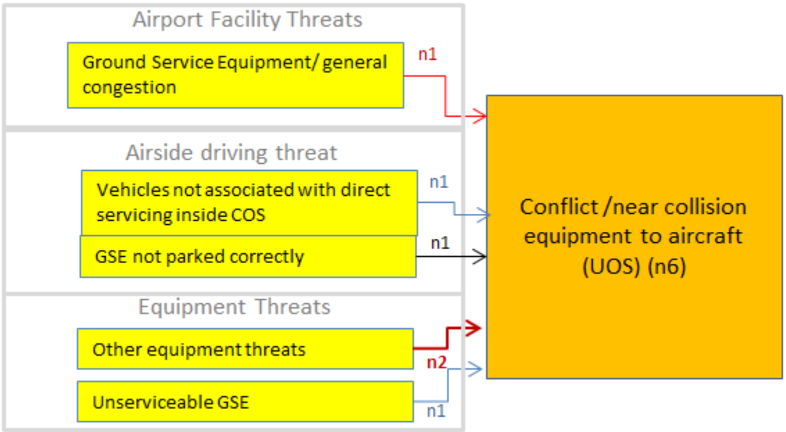


Figure 71: Threats associated with conflict/near collision with aircraft.

Figure 71 shows that two of these incidents were preceded by driving threats, where a vehicle not involved in the turnaround drove very close to the aircraft (inside the circle of safety zone) and another occasion on which the parking of ground servicing equipment brought it in conflict with the aircraft. Equipment usage threats, such as unserviceable GSE, also lead to a conflict or near collision. Airport congestion generally may increase situations where ground servicing equipment is driven too close to aircraft in the bay.

10.3.9 Threats associated with loading UOSs

There were six loading events associated with a preceding threat arising from the Freight and Baggage area or from an arriving irregular load. Addressing these threats should therefore reduce the opportunities for loading undesired states to occur.

Error! Reference source not found. (over) shows that four of these threats originated in Freight and Baggage departments. Although no variations were recorded in these observations, it would be expected that the ramp crew would manage these threats before they are loaded onto the aircraft. There were two loading events arising from arriving loads originating from other ports. Reducing these threats should reduce the overall opportunities for loading events generally.

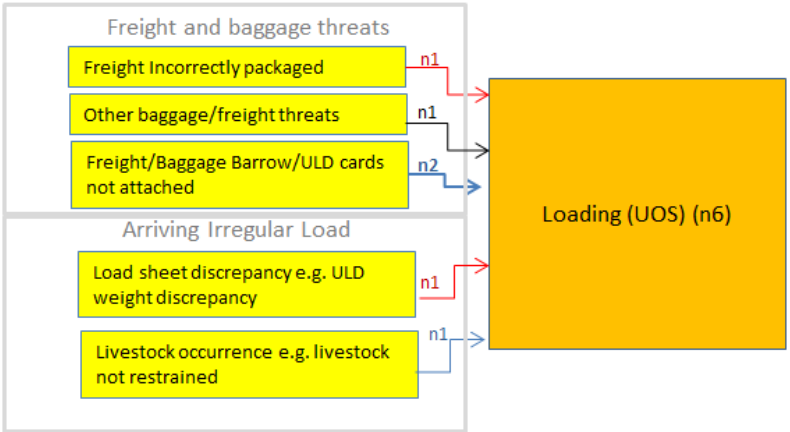


Figure 72: Threats associated with loading UOSs.

10.4 Relationships Between Threats Variations and Undesired States

It is also possible to use the NOM data to analyse relationships between threats, variations and UOS together. Any UOS can be assessed for the variations most likely to precede it and the threats that are present at the same time. These can also be represented within a combined diagram. In the example shown in Figure 73, the irregular loads events are preceded by loading variations, which are in turn associated with particular freight and baggage threats. Reducing these threats should reduce the likelihood of the associated loading variations, which in turn reduces the opportunity for loading incidents.

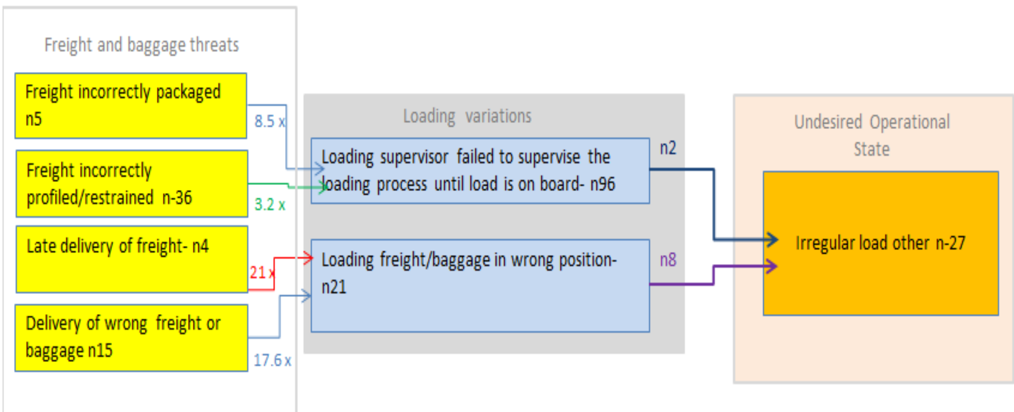


Figure 73: Relationships between threats variations associated with irregular loads.

There were 27 irregular load threats recorded. Ten of these were preceded by loading variations as shown in Figure 734. These variations are much more likely to occur when particular freight and baggage threats are present, such as delivery of the wrong freight, late delivery, or the incorrect profiling or package of freight. Addressing these threats in particular should reduce the odds of these specific loading variations and, therefore, the opportunity for a large proportion of the irregular loads to occur.

This process of working back from the UOS to identify the associated variations and threats could be repeated for all UOSs. Using the data in this way helps to identify ways in which the chain of associations might be broken to prevent particular UOSs from occurring. Designing interventions to address these chains of events should reduce undesired states overall.

10.5 Summary and Discussion of Data

10.5.1 Analysis using the NOM framework

Chapter 10 set out to investigate the relationships between threats variations and undesired states. Like LOSA, which uses the TEM framework, NOM used its own model of human performance to investigate which threats and variations to target for investigation. Rather than asking observers to decide which threats or errors are consequential during observations, as in LOSA, the NOM process simply asks observers to note threats, variations and outcomes so that associations can be analysed after the fact. The statistical analysis of associations between threats variations and UOSs provides a method for identifying significant relationships and associations between elements of the NOM model and, therefore, high-value targets for intervention.

The intention was to investigate which, out of all the possible combinations, were the noticeable associations or relationships between threats and errors that should be investigated. NOM also sought to identify the strengths of association through odds ratios.

This additional analysis highlights the relationships that are most influential and, therefore, the most appropriate targets for change.

Table 24, at the beginning of this chapter, provided a high-level overview of the odds of each major threat and variation category occurring together, and whether this association was statistically significant. There are 143 potential relationships that could arise from this analysis, but only 31 significant relationships were found. The analysis then drilled down to understand the nature of these relationships further, by investigating the associations between subcategories, to identify the individual threats and variations most likely to occur together. When considering all the potential relationship between the subcategories, there were 12,768 possibilities. The chances of any relationship being significant are therefore less than one in 10,000. However, many significant relationships were found, suggesting that, far from being random, the analysis produces lawful relationships between threats and variations in line with the proposed NOM model.

10.5.2 High impact threat categories

Table 24 indicates that some threat categories have more impact overall on the presence of variations such as freight and baggage threats and communication threats, which are discussed further below.

Freight and baggage. Freight and baggage threats were significantly associated with six major variation categories: documentation, driving, equipment usage, loading, read back and restraining variations. When considering the impact on documentation variations, Figure 62 demonstrates that freight and baggage threats increased the odds that the Load Instruction Report would be incomplete (n40), that crew would fail to refer to the paperwork (n39) and that hold operators' checklists would be incomplete or incorrect. Freight and baggage threats also increased the likelihood of driving variations such as not using the parking brake (n181), leaving vehicles unattended (n564) and incorrect reversing of vehicles (n14). These same threats also had a considerable impact on loading variations, as demonstrated in Figure 61, including the failure to check ULD serviceability (n31), failure to check for no fits (n74), failure to supervise the load (n96), failure to check the hold is fully discharged (n30) and failure to use absorption material for seafood and livestock (n6).

These results suggest that dealing with badly presented, profiled or mislabelled freight requires attention, time and resources of the crew that may detract from a wide range of other tasks. The relationship between these threats individually, or their collective effect on the performance of other loading tasks, might not have been appreciated without the benefit of NOM analysis. Reducing freight and baggage threats could, therefore, provide a high-value target for intervention that could have many previously unrealised benefits for improving the loading task overall.

Communication threats. A similar theme can be seen with communication threats. Communication threats appear to have a wide impact on five variation categories as shown in Table 24, including door, driving equipment usage, read back and unloading variations. Communication threats increased the likelihood that crew would fail to check the doors for damage (n85), carry out pre-driving inspections (n212), use a parking brake (n181) and would approach the aircraft before the 'thumbs up signal was given' (n55). As shown in Figure 58, communication threats also significantly increased the likelihood of unloading variations such as approaching the aircraft before the anti-collision beacon was turned off (n27) and failure to load according to the balance procedure (n5). Communication problems between Load Control and the Loading Supervisor dramatically increased the likelihood that the required information would not be provided in the read back (n168), by 246 times. Communication threats also increased the likelihood of equipment variations such as GSE touching the aircraft (n55), using the wrong equipment (n50), removing the belt loader whilst the operator is still in the hold (n36) and having multiple operators on the bridge of a JCPL while the rear platform is lowered.

These findings suggest that improving communication between all parties who interact with ramp crew could significantly reduce a wide range of variations experienced in this environment, suggesting that communication issues are a high-value target for intervention.

10.5.3 Low impact threat categories

Some subcategories had an unexpectedly low impact on the presence of variations. At the start of the stud, it had been anticipated, given the tight schedules, demand for on-time performance and the all weather conditions of ramp operations, that operational pressures, manpower and weather threats would have a high impact on ramp variations. However this was not found to be the case. In addition, it was found that although some threats were common, they had very little impact on variation. Examples of these unexpectedly low impact threat categories are discussed further below.

Operational pressures and manpower threats. At the start of the NOM implementation it was expected that operational pressures would have a significant impact on ramp operations, but this threat appears to have a very minor impact overall. It may be that observers were not aware of the operational pressure impacting the crew because it difficult to observe. Operational threats were rare and only seem to be associated with bay management variations, such as preparing the bay before the aircraft arrival. The availability of manpower to perform the turnaround was also not a significant issue. Larger crew sizes did not reduce the occurrence of threats or variations, and the major category of manpower threats was not associated with any variation category.

Weather threats. Weather threats were also expected to impact performance at the start of the study, given the wide variety of weather conditions encountered during aircraft turnarounds around the world. However, the threat was not significantly associated with any variation category. This may be a feature of the rules and procedures that prevent aircraft loading in extreme weather conditions such as high winds, hail or lightning. Further investigation is necessary before any conclusions can be drawn.

The circle of safety. Failure to perform the circle of safety was by far the most common threat, when other non-ramp crew failed to perform this procedure (n579, affecting 44.4 % of observations) and the most common variation recorded for ramp crew (n573, affecting 44% of variations). However, the analysis of associations in NOM revealed that failure to perform the circle of safety (as a threat) was not significantly associated with *any* variation. Despite the fact that there 573 variations of this type, it only appears once preceding an undesired state and even then occurs along with two other threat subcategories. Despite the very high levels of noncompliance, increased compliance may not lead to associated benefits based on the observations in this data set. The fact that the circle of safety is routinely ignored suggests that the procedure is either not seen as important or is not practical to comply with. The usefulness of this procedure as a control for preventing ground servicing coming into contact with the aircraft is therefore questionable, and alternative procedures and controls may be necessary.

This example also highlights the benefit of post hoc statistical analysis of NOM relationships when looking for intervention targets, rather than focusing only on high frequency threats and variations. Typically, high frequency threats or errors in LOSA become key targets for intervention. However, NOM analysis would suggest that this is not always advisable, as high frequency does not necessarily mean high impact. The high frequency of circle of safety threats and variations provides a good example of this.

10.5.4 Targets for change – improving organisational defences

The relationships between threats, variations and undesired operational states highlight some key areas where interventions could be targeted. Some of these improvements are targeted at the ramp crew’s ability to manage threats and variations, whereas others target the systems level to improve organisational defences. The relationships discussed in this chapter suggest that addressing these issues in particular would have the most impact on reducing risks on the ramp.

Improving the airport facility. Figure 54 shows how unclear or missing ramp signage dramatically increases (12.3-fold) the likelihood of incorrect equipment staging by ramp crew. This was related to GSE congestion generally, which occurred 77 times and was associated with incorrect staging (n41) and the failure to perform basic supervisory checks such as checking door sills for damage before opening (n85). This would suggest that addressing some of the features of the airport facilities would reduce the incidents of this type of variation. For example, the lack of parking and staging areas, particularly in busy ports, appears to be associated with a number of more serious issues. The extent to which these threats increase issues such as driving under the wing and the potential for aircraft damage was not previously acknowledged and may not have been realised without evidence from NOM statistical analysis. With airports predicted to get busier, and space at a premium, NOM data highlights the design of space at airports may have important ramifications for the safe movement of equipment and aircraft on the ramp in the future.

Improving access to equipment and materials. It can be expected that if personnel are not provided with the right equipment and materials, they will not be used. Although not a surprise, the NOM data provides evidence to demonstrate the effects of this relationship. Figure **60**61 shows how unserviceable components and equipment, or equipment left in an inappropriate condition, increases the likelihood of equipment usage variations. Similarly, we can see from Figure 61 that if absorption materials are not provided there will be an increase in variations and UOS where absorption materials were not used. Failure to use the right equipment and material can lead to aircraft damage. Use of the incorrect GSE can lead to damage to aircraft doors and sills, or to aircraft holds. Transporting seafood or livestock without absorption material, for example, can lead to leakage of fluids (such as seawater) that cause damage and corrosion and shorten the life of the aircraft components. Theoretically, these variations should be easy to rectify if the appropriate materials and equipment can be provided at the time they are needed.

Improving compliance on the ramp. The results of this study indicate that there are a high number of variations occurring with standard operating procedures on the ramp during normal operations. This suggests that work as performed varies from the procedures in some way much of the time. The causes of variations requires further research, but a first start would be to improve rules that are impractical or ambiguous, such as the circle of safety or the rules for the use of refuelling cones. A clear, workable ruleset is important for preserving the integrity of standard operating procedures and reducing the gap between how we say work should be done and how it is actually performed.

NOM data provides clues and insights into where and when rule compliance was a problem, which can then be further investigated to understand why people vary from the procedures. We can see in Figure 52, for example, that there is an association between non-ramp crew approaching the aircraft too soon (recorded as a threat) and ramp crew approaching the aircraft too soon (recorded as a variation) prior to the beacon being extinguished. It may be, for example, that once one person approaches the aircraft, and it encourages others to think it is safe to do so. In another example, Figure 556 indicates that when the threat of 'marshaller is not present' is observed there is an 8.3-fold increase in ramp crew approaching the aircraft while the anti-collision beacon is still activated. The influence of rule breaking and group norms and supervision, therefore, may be useful starting points for investigating the influences on rule compliance across the ramp environment.

Compliance with rules is more likely when the risks the rules are designed to control are understood (Mason 1997). NOM data can be used to provide evidence where the risks associated with rule violations may not be known. Figure 55, for example, indicates a link between the threat ‘towing more than allowable rolling stock’ and the variation of drivers failing to come to a full stop. Increasing awareness of how increased weight of rolling stock can affect braking distances, for example, may increase risk perception for this rule. Increasing risk perception using evidence from NOM may improve compliance generally.

Improving supervision of the loading task. Supervision is intended to provide a control mechanism for recovering from things that go wrong during unloading and loading activity. Figure 54 provides one of several examples that suggest a relationship between dealing with threats and the omission of basic supervisory tasks such as checking the door sills. Figure 578, for example, indicates that the failure of the engineer to prepare the bay prior to the aircraft’s arrival (n10) also increases the likelihood of this check being skipped, by 6.3 times. Some threats also affected supervisory checks in the loading tasks. The Loading Supervisor visually checking door sills, locks guides and restraining stanchions before releasing the aircraft for flight (n254) was 7.8 less likely to happen if the crew were faced with the threat of incorrect or incomplete documentation. This threat also led to a 17-fold increase in the load being incorrectly restrained. This suggests that having to resolve documentation issues during the turnaround reduces the time or available resources to perform the normal checks of the load and restraint prior to take off.

The standard operating procedures state that the Loading Supervisors is responsible for a number of other tasks that may direct his attention away from the loading task during the turnaround. The Loading Supervisor failing to supervise was recorded 96 times. Figure 61 indicates this was most likely to be preceded by freight arriving incorrectly packaged (n5), which increased the likelihood of occurrence by 8.5 times. Similarly, the load arriving incorrectly profiled or restrained (n36) was associated with a 3.2-fold increase in the failure to supervise. Problems with arriving freight, as well as over-stacked barrows (n72), were also associated with an 8.2-fold increase in the failure to check the hold was fully discharged, as shown in Figure 61.

Failure to supervise variations were also involved in the occurrence of loading undesired operational states (UOSs). Figure 73 indicates there were two occasions where failure to supervise the entire load contributed to irregular loads. Addressing these freight and baggage issues might lead to an unexpected improvement in supervision and monitoring tasks generally, which could have dramatic impacts on performance for a wide range of loading risks.

Several of the figures in this chapter suggest a relationship between becoming distracted by dealing with threats, and an increase in variations. This lends some weight to Heinrich's original argument that the presence of minor unsafe acts and conditions increases the likelihood of more serious incidents occurring. Without the NOM data, addressing these apparently minor threats might not have been a priority; however, with NOM data it is possible to see which threats offer high-value targets for improving safety.

10.5.5 Implications for the NOM model

The data collected in this ramp NOM implementation also provides an opportunity to test the NOM model proposed in Chapter 7, to assess whether the assumptions of the model were supported by the data. This section discusses how and where the data supports the model and identifies where amendments may be necessary.

NOM as a measure of current performance in Normal Operations

The data from this study provided a bassline measure of how tasks are performed in line with standard operating procedures. Variations were common, suggesting there is a significant gap between work as planned and work as performed. There were over 3137 variations recorded across the 1302 observations conducted. Most observations (93.6%) recorded at least one variation from standard operating procedures, with the maximum recorded in any one observation being 17. NOM was able to provide a measure of current performance as a baseline from which future performance improvements targets can be set and measured.

Normal, abnormal and emergency modes. The NOM model in Figure 15 makes the assumption that, most of the time, threats and variations will be managed within safe limits as part of normal operations. The data collected in this NOM observation broadly reflected this assumption. For example, most of the observations (87.3%) were of normal operations, with a small proportion (12%) reflecting circumstances when operations became abnormal. There was only one observation (0.7%) where a recordable accident was observed (Emergency operation).

In a change from LOSA, which records links between threats and errors during the observations, the NOM observation methodology identified statistically significant associations between elements of the model using post hoc statistical analyses. Inherent in the NOM model was the assumption that this data could help to identify potential accident trajectories. This assumption has been supported by the data. The associations between threats, variations and Undesired Operational States discussed in this chapter identify potential pathways via which an operation can shift from normal to abnormal or emergency operations. The method was also able to identify some limitations of the organisational system and defences.

In addition, the NOM model assumed that the identification of 'management' behaviours would also highlight the ways in which crews could exercise resilience, to manage or return the operation to normal, safe limits. The pathways are shown in green lines in **Error! Reference source not found.**Figure 15 and Figure 74. The data from this study offers some support to this concept; however, it was expected that the number of threats and variations that were managed would be much higher. Unlike previous LOSA implementations, few threats (11%), variations (1.4%) or UOSs (6.1%) were observed to be managed in this ramp NOM implementation. This may be a feature of the ramp environment, where there are multiple different groups converging in the same space, often under time pressure, meaning threats and variations to SOPs are tolerated rather than managed. Or it could be that the wider context of the observational methodology on the ramp made it more difficult to detect management of threats and variations, although this seems unlikely to fully explain the very large number of unmanaged threats and variations. Further investigation is necessary to understand why the management rates were lower than expected.

Threats. The data from NOM suggested that threats were common in normal operations (n2289) with almost all (97.7%) of observations involving at least one threat. As suggested by the NOM model, the observations showed that not all threats need to be managed to remain inconsequential. The data from this NOM implementation, however, suggests that threats can lead directly to an Undesired Operational State without being poorly managed, or being associated with an intervening variation. This represents a difference from the LOSA TEM model described in Figure 7, and from the NOM model described in Figure 15. There were 22 UOSs associated directly with a preceding threat. Almost all were the result of poor driving practices of non-ramp crew near the aircraft being serviced. These UOS events included non-ramp crew vehicles coming into conflict with aircraft (n8), with structures and parked equipment (n6), with personnel (n3) or with each other (n3). Two other events involved dead or injured livestock arriving as irregular loads. As such, they were threats that resulted directly in undesired states without any possibility for ramp crew intervention. The proposed NOM model in Figure 15 has been amended to allow threats to lead directly to a UOS. The amended NOM Model shown in Figure 74 represents this as a red dotted line and arrow.

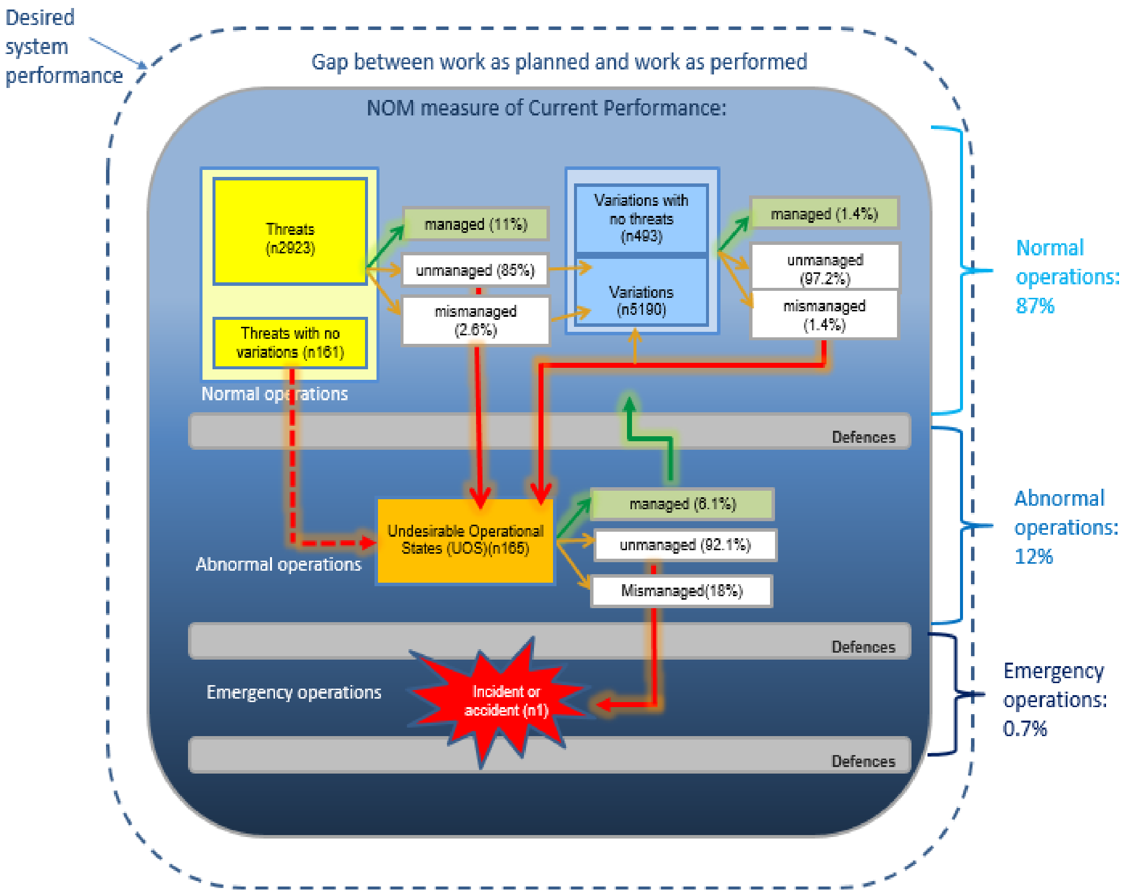


Figure 74: Amended NOM model.

Variations. In the TEM model shown in Figure 74, crew errors arise from poorly managed threats. In the NOM model, errors were redefined as variations that also occur as the result of poorly managed threats, but in a departure from TEM they can also occur spontaneously, in the absence of threats. This was supported by the data from this NOM implementation. As shown in Figure 74, there were 160 observations that recoded a variation but involved no threat. This figure may include some occasions where a threat was missed (not observed/not recorded) by the observer. The comparatively large number, however, would seem to support the concept that variations can arise spontaneously as suggested by Klinect (1996).

Undesired states. In the TEM model, undesired states arise from mismanaged threats and errors. In this NOM implementation, undesired states also arose from mismanaged threats (n41) and variations (102). In a change from TEM, however, the data from NOM also suggests that UOSs arose from threats outside of the ramp crew's control or influence (n22), as discussed above and shown as a dotted lines in in Figure 74. This finding suggests that, in this environment, NOM was able to identify opportunities for accidents to occur that were not under the direct control of the crew being observed. This may be indicative of the ramp environment, where the actions of many interacting personnel have the potential to cause undesired states more often.

Organisational defences. Although not coded during the observations, the NOM data was able to identify areas where the organisational system was deficient. Issues with the airport facilities and equipment, as well as poorly defined rules and procedures, increased the opportunities for variations and UOSs to occur. The NOM data also allowed identification of high-value targets for improving organisational defences, such as improving training and managing threats originating from other service providers such as Freight and Baggage.

Implications for the model. The NOM model has been broadly supported by the data from this NOM implementation. Further trials of the method in other environments would be helpful to test its suitability in other operational domains. In this context, the model suggests that we should be looking for ways to reduce variation from work as planned by improving the way tasks are prescribed, increasing compliance, increasing resilient management behaviours, improving organisational systems and defences and disrupting the links between associated threats variations and UOSs to block potential accident trajectories and prevent the next accident from occurring.

10.6 Summary and Conclusion

Any one of the combinations could have occurred by chance; of those threats and variations which were strongly associated, some may have been expected but others would not have been identified without the benefit of statistical analysis of NOM data. The data provides a measure of how current performance varies from the tasks as prescribed. It also provides an evidence-based approach to identifying which threats and variations need to change in order to produce safer outcomes. If variations increase the opportunities or likelihood of undesired states and safety incidents, then it makes sense that the industry would want to address the precursors of such situations. Data from NOM studies helps to identify, statistically, where these precursors and associations are, so they can be managed. The data also indicates where improvements are needed at a systems level to improve organisational defences. Armed with this information, the organisation can make informed choices about where best to intervene to have the most impact on safety.

The analysis so far, therefore, has only begun to scratch the surface of what could be analysed. Whether these associations represent causes and effect relationships requires further investigation and analysis, in order to validate the most effective solutions. However, we can see even from this limited analysis that NOM data is able to provide new insights that are not available from existing data sources. The next chapter assesses the NOM method against the evaluation criteria outlined in Chapter 6.

Chapter 11 Assessment of NOM against the Evaluation Framework

Chapter 6 outlined a framework for evaluating the NOM methodology based on a number of criteria. These criteria were first applied to evaluate the strengths and weaknesses of LOSA and to define the ideal features of a normal operations methodology for ground operations. The framework outlined seven criteria:

1. Comprehensiveness
2. Consistency
3. Usefulness
4. Use of resources
5. Documentability
6. Acceptability
7. Validity.

This chapter will evaluate how the resulting NOM methodology has met the development needs identified in Chapter 6 and consider how well it meets each of the seven criteria described in the evaluation framework outlined in Chapter 6.

11.1 Comprehensiveness

Comprehensiveness: *The completeness with which an observation tool can describe the full range of human behaviour observed as well as observable factors in the environment which influence those behaviours – rated on a scale of 1 (poor) 2 (moderate) or 3 (good/high).*

Evaluation of NOM for comprehensiveness

The NOM provided a comprehensive range of domain-specific codes based on an extensive development and refinement process, as described in Chapter 8. This process helped ensure the resulting taxonomy was inclusive of all aspects of ground crew performance encountered during the observations.

Threat categories covered a wide range of performance-shaping factors identified by subject matter experts. Variations codes were developed based on a comprehensive review of rules and a task analysis and on a HAZOP study of all the ways in which humans could vary from the procedure described, as outlined in Chapter 8. The use of 'other' categories for threats and errors was rare; only 6 per cent of threats, and 1.56 per cent of variations, were coded as 'other', indicating that the codes were necessary and sufficient for observers to describe what they saw more than 94 per cent of the time.

The new method was able to identify both successful and unsuccessful behaviours in terms of their outcomes using the managed, unmanaged and mismanaged codes. Furthermore, the narratives provided a rich contextual background to interpret how situations were experienced by the crew at that time. The data provided by NOM was thus very comprehensive. NOM has therefore been rated as 3 (high) for comprehensiveness based on the criteria above.

11.2 Consistency

Consistency: *The degree to which the technique is structured to achieve reliability so that different assessors to describe and categorise the same observable phenomenon in the same way – rated on a scale of 1 (poor) 2 (moderate) or 3 (good/high).*

Evaluation of NOM for consistency

The changes to the error classification described in Chapter 7 and the coding taxonomy helped to reduce ambiguity. Terms were altered to describe only observable phenomena based on variations from standard operating procedures, which reduced the need for interpretation and subjectivity when assigning codes, and should therefore have improved consistency in their application. Extensive training was given in coding observations. The NOM training program was extended from the recommended LOSA training program to allow more practice using real examples and staged videos. Observers were not permitted to begin observations on the ramp until a standard of coding accuracy was achieved, as described in Chapter 8. NOM also retained the data verification process or 'wash' to increase the consistency of codes assigned. This allowed codes to be assigned via a team of observers where differences of opinion arose. This process avoids many of the problems associated with consistency between coders since inconsistencies are resolved via consensus.

Although inter-rater reliability was not formally measured as part of this study, data on each observer was collected for comparison. Where data from any particular observer varied from the group, feedback was given to support the ongoing calibration of the observation team. However, ongoing monitoring and calibration of observers is needed, particularly as the observer pool expands or changes in the future. Modifying coding and training helped to ensure that different assessors can describe and categorise the same observable phenomenon in the same way. According to the criteria above, the NOM method would therefore rate a 3 in terms of consistency.

11.3 Usefulness

Usefulness: *Priority given to generating useful improvement strategies and the perceived usefulness of the tool and its outcomes to stakeholders – rated on a scale of 1 (poor) 2 (moderate) or 3 (good/high).*

Evaluation of NOM for usefulness

NOM's usefulness depends on whether or not it is perceived as useful by all stakeholders and whether or not it can generate useful improvement strategies for ground handling organisations. Some stakeholders in the academic community had doubted LOSA's usefulness due to concerns regarding the method's validity (Decker 2003; Dekker and Hollnagel 2001), as discussed in Chapter 6. NOM has attempted to address some of these concerns and the evaluation of NOM's improved validity is discussed further in 11.7 below. NOM's perceived usefulness to industry will be based on how well it is perceived as providing the benefits attributed to LOSA and whether or not it is perceived as being able to provide useful improvement strategies. Both are discussed with reference to NOM below.

Providing useful improvement strategies. NOM data was able to provide high-value targets for intervention that, if addressed, would reduce the likelihood of variations and the opportunities for undesired states to occur. The usefulness of interventions based on this data was not evaluated as part of the current research. However, several of the groups involved reported that the data has been useful in assessing a baseline measure of actual vs planned performance; improving the quality and usability of procedures; identifying the strengths and weaknesses of the operation; addressing problems arising between different departments; targeting training development needs; setting future performance targets; and developing improvement interventions. Initial results suggest the data was useful to the participants of this study. Further research would be needed to measure the perceived usefulness of these outcomes to industry

Providing similar perceived benefits to LOSA. NOM's usefulness is also likely to be dependent on whether it is perceived to provide similar benefits to LOSA whilst hopefully also providing some improvements. To estimate this, NOM can be evaluated against the 9 LOSA benefits described by ICAO (2006) and discussed in chapters 5 and 6. The potential for NOM to provide similar benefits in each of these nine areas is discussed below.

(1) Identifying threats in the operating environment: NOM was able to identify 2.63 threats per observation. It was also able to identify associations between threats and demographic and operational data. This makes it possible to identify not just threat frequencies, but associations between threats and other conditions such as team size, port type, geographic location, specific working arrangements and bay layout, for example. The benefits included being able to identify problematic locations (such as localised equipment shortages) as well as organisational issues occurring across interfaces, such as catering, load

control, freight etc. The data indicates how threats originating in one area, such as Freight and Baggage, for example, can contribute to problems such as irregular loads. NOM post hoc statistical analysis was able to demonstrate empirically which threats were associated with which variations and Undesired Operational States, as discussed in Chapter 10.

(2) Identifying errors in the operations: NOM did not identify errors according to the original LOSA taxonomy but instead identified 'variations' where the crew's actions varied (for whatever reason) from the standard operating procedures or work rules. NOM was successful at identifying variations and coding a total number of 5190 variations and 2.62 per operation (SD 2.62) per observation. In fact, only six per cent of observations recorded no variations at all. This suggests that crew were not adjusting their behaviour in the presence of the observer, or that, if they were, they were not able to prevent making mistakes or variations from standard operating procedures even in the presence of the observer. NOM helped the organisation understand which variations were occurring most often, such as airside driving variations, and even specific locations where variations were more likely. NOM post hoc statistical analysis was able to demonstrate empirically which variations were associated with Undesired Operational States, as described in Chapter 10.

(3) Assessing the degree of transference of (CRM/ NTS) training to the line: The current research has not tested the degree of transference of training from the classroom to the operational environment, but the method could be used to do this in the same way LSOA does. At the time the NOM observations began, NTS training was not standard practice for ramp personnel. NOM was, however, able to identify which behaviours were most likely to result in successful or unsuccessful outcomes, based on evidence from the observations. This can help identify specific targets future training. Examples from the narratives can also be used to provide realistic case studies and training scenarios.

(4) Checking the quality and usability of procedures: NOM was able to identify many problems with the quality and usability of procedures, as discussed in chapters 9 and 10. NOM identified procedures with high rates of non-compliance, such as the circle of safety (573 recorded variations) and identified problematic procedures, such as protecting refuelling lines with traffic cones, where the rule itself may be ambiguous, as described in Chapter 9. Port reports also indicated where local procedures differ from official ramp manuals, such as in procedures for towing rolling stock. This may have contributed to differences in behaviour observed at some ports, but also provided an opportunity to clarify and improve the consistency of rules and their application across ports. One carrier involved in the study reported that many simple day-to-day procedures have been adapted as a result of data from the observations, which has helped to streamline ramp practices and improve staff engagement (Ma et al. 2011).

(5) Identifying design problems in the human machine interface: NOM was able to identify design problems with equipment and the working environment more generally. There were, for example, 447 equipment threats recorded, such as a high number of damaged Unit Load Devices (ULDs). As a result of the NOM study, a separate investigation was launched in an associated freight and baggage organisation into the design of the ULDs and the nature of ULD failures. In addition, the ramp manual was updated to provide additional tools to support staff in deciding when to reject a ULD on the grounds of damage. NOM also identified problems in the design of aircraft and airport facilities. There were, for example, 148 airport facility threats, some of which were found to contribute to rates of congestion and associated variations, as discussed in Chapter 10. Comparison by location also indicated that these threats were higher on some bays than others, suggesting the bay layout or its specific features may be affecting ramp team performance on these bays. Analysis and comparisons of bay features and performance at different locations could be used to identify optimal features of bay layouts to

inform future airport design. NOM data has also been used to inform better equipment design, with manufacturers using the data to guide the design of new equipment to improve human performance on the ramp. (Ma et al. 2011, p. 6).

(6) Understanding pilots' short cuts and workarounds: NOM was designed to record short cuts and workarounds as variations irrespective of the cause of noncompliance. It was, however, possible to determine which rules were routinely complied with, as well as the frequency, location and circumstances under which all the variations occurred, which may include shortcuts. The narratives also provided rich context and insight as to why some rules were varied more than others. For example, data from NOM indicated that leaving tugs running while unattended was associated with having to repeatedly jump-start old ground servicing equipment. Similarly, driving under the wing or too close to a parked aircraft was associated with insufficient parking and staging, as discussed in chapters 9 and 10. NOM data helps to understand where to target change to address noncompliance.

(7) Assessing safety margins: The NOM method was able to capture 165 Undesired Operational States (UOSs). If the large data set can be considered representative of the operational network, it is possible to extrapolate how often these UOSs would be experienced in normal operations generally, as a measure of current safety margins or risk exposure, as well as where these are more likely to occur. Understanding an operation's potential exposure to undesired states in normal operations provides a measure of safety that may be helpful for driving continuous improvement in safety performance.

(8) Providing a baseline for organisational change: NOM data provides a measure of safety performance as outlined in area 7, above, as a baseline for organisational change. The data from NOM also provides high-value targets for developing safety interventions, as discussed in Chapter 10. Repeating NOM observations in the future could provide an

opportunity to evaluate whether interventions based on NOM data and targets produce the desired changes; however, this was not assessed as part of the current research.

(9) Providing a rationale for allocation of resources: As discussed previously, NOM data indicates high-value targets for the allocation of resource. For example, NOM indicated where there were particular problems associated with faulty equipment and airport facilities. Allocating resources to address these problems could reduce the odds of associated variations and undesired states. NOM also provided data about issues contributing to operational delays. One airline reported that allocating resources to managing the issues reduced delays and improved scores of customer satisfaction, as well as improving safety (Ma et al. 2011, p. 6).

Based on the ability of NOM to provide similar benefits to LOSA and identify useful targets for intervention, as discussed above, NOM scores a 3 (high) in terms of its usefulness for generating improvement strategies and its perceived usefulness to industry.

11.4 Use of Resources

Use of Resources: *The cost of implementing the tool in terms of training expertise, time and resources – rated 1 (poor) suggesting that burden to industry is high, 2 suggesting a moderate resource burden or 3 (good) to indicate low use of time and resources.*

Evaluation of NOM's use of resources costs and benefits:

Setting up a NOM program, like LOSA, takes considerable up-front investment in terms of time and resources, including training the observers, collection of over 1300 observations and analysis of data When compared with LOSA, NOM may make some savings through the use of more prescriptive coding, which may reduce the time and resources required for coding and data verification. Whether the use of resources is acceptable depends on the perceived value of the program. Establishing the value of such programs could involve tracking other cost benefits, such as reduction in costs associated with delays, aircraft damage or fines associated with regulatory breaches; however, this was beyond the scope of the current study. More empirical evaluation is needed to determine the cost/benefit ratio of NOM to justify the investment. Future research should also consider how the program resources could be reduced. Therefore, based on the definition above, NOM can only be ascribed an interim rating of (1) for its use of resources; this could be re-evaluated in the future.

11.5 Acceptability

Acceptability: *The degree to which all stakeholders can accept the underlying model of human performance and, therefore, the results generated by the method – rated as 1 (poor) 2 (moderate) or 3 (good).*

Evaluation of NOM for acceptability

The NOM model proposed in Chapter 7, and further discussed in Chapter 10, is based on the Threat and Error Management (TEM) model first proposed by Klinect, Wilhelm and Helmreich (1999), which has been widely accepted by the aviation industry. However, TEM was not without its critics and a number of concerns about the model and its validity were discussed in Chapter 5. To address these concerns, the TEM model was simplified and amended in terms of the taxonomy and relationships depicted within the model, as discussed in Chapter 7. Most of the original elements of the TEM model are still present in the NOM model, but are modified to describe only observable phenomena, as befits an observational methodology. It is hoped that, based on these changes NOM, as an underlying model of human performance, would be as acceptable as TEM to industry, whilst offering a more acceptable model for those who have previously raised methodological concerns with regard to TEM (Hollnagel and Amalberti 2001; Dekker 2003; Dekker and Hollnagel 2004). The results from this implementation broadly support the NOM model, as discussed in Chapter 10; however, more research is necessary to establish its validity. Based these improvements, it is argued that NOM offers an improved model of human performance compared with LOSA, and is therefore rated as ‘good’ (3) in terms of its acceptability to stakeholders.

11.6 Documentability

Documentability: *How well the method and results are documented to allow periodic audits to provide a baseline of performance and drive continuous improvement – rated as 1 (poor) 2 (moderate) or 3 (good).*

Evaluation of NOM for documentability and auditability.

NOM as a method has been fully documented in Chapter 7 for the purposes of its application in any industry. Its adaption and application in ground handling has also been fully documented in Chapter 8, in sufficient detail to allow others to develop similar ground handling NOM programs. Like LOSA, NOM is a form of audit and, as such, it provides baseline measures for driving continuous improvement. The data from NOM can be documented and made available for purposes such as investigation, training, design or input to developing systems or developing procedures. Periodic monitoring can demonstrate whether interventions area achieving their desired aims and allow for continuous improvement and feedback. Like LOSA, NOM therefore rates as a 3, as it has high documentability based on the criteria described above.

11.7 Validity

Theoretical validity: *The validity of the underlying theory or construct of human performance on which the tool is based as well as the validity of the data collection methodology employed – rated as 1 (poor), 2 (moderate) or 3 (good/high).*

Evaluation of NOM for theoretical validity

Chapter 6 describes how NOM attempted to retain features of LOSA that were considered strengths and to adapt features of the LOSA model and data collection methods where weaknesses were identified to improve the validity of the data collected.

The LOSA and TEM observation method and codes have been modified in the new NOM model and method with the aim of improving theoretical validity in terms of the method and the data generated. In an important difference from LOSA, the NOM approach is restricted to recording only observable phenomena, which do not require interpretation on the part of observers. For example, in NOM observers are not required to make assumptions about what an operator intended to do, or the cognitive mechanisms of failure when the behaviour observed was not as expected. Replacing error with a neutral descriptor of 'variation' from a procedure removes interpretation of causes or confusion with their outcomes, which can occur in LOSA, as discussed in Chapter 5. In NOM, observers must only record the presence of threats, behaviour deviating from standard operating procedures, and the outcome.

In LOSA, observers are asked to determine whether threats and errors are 'consequential', that is, leading to other errors or undesired aircraft states. NOM relies instead on multivariate analysis of the data to identify statistically significant associations between threats, variations and undesired states in order to identify which threats may make certain variations more likely.

Some aspects of the NOM method were deliberately replicated from LOSA in order to retain features of the methodology that strengthen validity. For example, LOSA uses large sample sizes to ensure the data is representative of normal operations. So, too, in this NOM implementation a large sample of over 1300 observations from 10 different countries was collected to create a representative sample of normal operations. In addition, many of LOSA's safeguards were retained to ensure subjects behaved naturally during observations, to ensure the method was capturing normal behaviours. This included using trusted peer observers and protocols for protecting confidentiality, for example.

NOM has attempted to address many of the concerns regarding LOSA's methodological validity in order to offer an improved tool with more reliable data upon which to base interventions for improvement. Based on these improvements, NOM would be rated as having good (3) theoretical validity; however, more research would be needed to test this empirically.

11.8 Summary of Evaluation against the Evaluation Framework

Using the criteria established in the evaluation framework in Chapter 6, NOM scores favourably in most areas. When compared with the original LOSA method, NOM could be said to offer improvements. In the area of comprehensiveness, NOM offers a more comprehensive taxonomy of codes. In terms of consistency, the simplified NOM model and codes offer some improvements to LOSA and TEM in its suitability as an observation method. Where possible, the strengths of the original method, such as large sample sizes and confidentiality, have been retained, and where weakness in the methodology have been identified, NOM has attempted to offer alternative approaches. The adapted NOM model, coding and analysis, as well as the use of multivariate statistical analysis to identify associations between elements of the model, allow for improved reliability of the data collected. This also allows conclusions to be drawn about where to target safety improvements. Based on the discussion above, Table 25 compares scores for each criterion in the evaluation framework for LOSA as described in Chapter 6 and for NOM based on the discussion above.

Table 25: Comparison of NOM and LOSA against evaluation criteria.

Evaluation Framework Criteria	LOSA evaluation scores (from Chapter 6)	NOM
1.Comprehensiveness	2	3
2.Consistency	2	3
3.Usefulness	2	3
4.Use of Resources	1	1
5.Acceptability	3	3

6.Documentability	3	3
7.Validity	2	3

In summary, when considered in terms of the evaluation criteria developed in Chapter 6, NOM scores higher than or equal to LOSA across all criteria, and therefore can be said to have met the development objectives at the end of Chapter 6. The next chapter provides a discussion of the research overall in terms of its original aims, its implications for ground safety and beyond.

Chapter 12 Discussion

12.1 Satisfaction of NOM Objectives

This research began with an observation that the lack of suitable methods for understanding human and safety performance in aviation ground safety may be impeding safety progress. The stagnation in safety improvements (Passenier 2015) discussed in Chapter 2, combined with increasing air traffic through already congested airports (IATA 2014), has made this a pressing concern for the industry. New insights are needed to understand why ground safety is not improving at the same rate as safety in the rest of the aviation industry. Finding new leading indicators is now necessary to drive improvements. The research therefore set out to develop a new method for measuring human and safety performance in ground operations. Normal Operations Monitoring is put forward as an innovative solution to meet this need of ground safety, but the result has been a method that can be applied anywhere that operational tasks can be observed.

NOM has been developed based on influential ideas from previous accident causation models discussed in Chapter 3. The Threat and Error Management model (TEM) (Klinect et al. 1999) was selected as one of the most suitable models for the basis of the current research. Methods for collecting data are discussed in Chapter 4, and observation, and the LOSA methodology (Klinect et al. 2003), with its aviation relevance, was selected as the most promising data-collection tool for adaptation in the current study. When scrutinised more closely, however, a number of issues were highlighted with regard to both the TEM model of human performance and the way in which LOSA collects data through observations (Dekker 2003; Dekker and Hollnagel 2004). Since its inception in 1999, LOSA has remained largely unchanged, despite new theories and research in the field of safety science. This research therefore set out to build on the strengths of LOSA while addressing the potential weaknesses, and incorporating new ideas where possible, to create an improved method for monitoring normal operations.

An evaluation framework was developed based on criteria proposed by Kirwan (1993) to help define the ideal characteristics of a new tool, as discussed in Chapter 6. This provides an objective standard to evaluate whether the adapted approach offers any improvement from the original LOSA method. NOM attempts to build on the credibility and face validity of LOSA by retaining the 'LOSA Operating Characteristics' (ICAO 2002) for data protection, confidentiality and representative sampling of normal operations, as described in Chapter 5. To overcome some of the perceived limitations of LOSA and the TEM model, NOM changes the original TEM terminology, approach to observations and analysis, as described in Chapter 7.

NOM includes a new model of human performance, based on TEM but incorporating elements from a number of models, such as Reason's concepts of organisational defences and the concept of managing behaviours as opportunities for organisational resilience. NOM focuses on analysing the gap or variance between work as imagined and work as performed (Hollnagel 2011). Neutral and unambiguous codes were developed, including substituting 'variations' from standard operating procedures for the 'error' category to focus instead on the 'variance' between work as planned and work as performed.

The NOM observation methodology is restricted to recoding only observable phenomena in the model, such as threats, variations, undesired operational states and their management. Objectivity is improved by the use of post hoc statistical analysis to investigate associations between elements of the model, rather than relying on observers to make assumptions or links between causes and effects.

The resulting NOM model and tools are adapted in this study for application and testing in the ground handling environment, but the study is also a case study of how they might be used in other similar operations. This case study was extensive, with over 1300 observations conducted, coded and analysed to provide insights into the threats faced by ramp crews, the ways in which their behaviours vary from standard operating procedures, and opportunities for undesirable operating states to occur. The NOM method also analyses all the potential relationships between observed threats and variations, and calculates the odds of those associations occurring by chance. The results highlight high-value targets for intervention, such as the particular threats and variations that, if addressed, could have a dramatic impact on the likelihood of Undesired Operational States occurring. The data from narratives, observer issues and port reports provides additional context. This qualitative information is also useful for highlighting specific problems with organisational defences, such as ambiguous or impractical rules, issues with the airport facilities and problems between different organisations and that interface with ramp operations. The implications of these results for the ground handling industry, and for the NOM method's ability to support high-hazard industries generally, are discussed further below.

12.2 Implications for Ground Safety

This application of NOM provides an opportunity to test the method and tools, but also to collect a large body of data from normal ground operations. The results clearly identify the variance or gap between work as planned and work as performed in the ground safety environment. More than just identifying the gap, the results provide insights into how this variance contributes to safety outcomes. The data provides new information about which particular behaviours are associated with safe or unsafe outcomes, for example. The associations between threats, variations and UOSs indicate clear targets for improvement in risk controls and defenses as well as human performance. This type of analysis also identifies the strength of those associations. In doing so, NOM is able to identify where effort would be best directed to influence outcomes on the ramp, as discussed in Chapter 10.

Chapter 10 identifies where improvements could be made in the organisational systems and defences. The results suggest, for example, that addressing issues identified with the design layout and maintenance of airport facilities, as well as the provision of appropriate materials and equipment, could have a significant impact on safety outcomes.

Unlike other previous LOSA adaptations, few ‘managing’ behaviours are observed when compared to data from similar observations in the cockpit cabin and rail environment. This may due to a number of key differences, such as a lack of training, motivation or opportunity for ramp crew. Ramp crew are not routinely trained in threat and error management, so may not be aware of the behaviours most appropriate to manage them. The results also suggested some specific targets for training, such as improving airport driving behaviours, the importance of improving supervision and how to manage issues between departments and service providers more effectively.

Lack of training, however, is unlikely to be the only reason the ramp exhibited low management rates. There may be other limitations that affect the way ramp crew can respond. This data suggests that there are some threats that happen near the ramp crew being observed for which there is no management action available, such as two non-ramp vehicles coming into conflict with each other nearby. Ramp crew may also be limited in their ability to influence risks originating between teams of other service providers , which would be better addressed at the system level. The time available to the ramp crew during the aircraft turnaround (typically between 20 and 40 minutes) also limits the range of available responses.

Overall, the NOM study suggests that – assuming the data is representative of normal operations on the ramp – undesired states are occurring in around 12% of all turnarounds. If each one of these is an opportunity for an incident to occur, then addressing these potential accident trajectories should be a priority. NOM data provides novel insights that inform high-value targets for improving safety and human performance on the ramp.

12.3 Assessment of NOM against Evaluation Framework

The evaluation of NOM using the framework outlined in Chapter 7 concludes that most of the development objectives for the new tool have been met. The neutral terminology and coding of variations from standard operating procedures has reduced the need for observers to interpret cognitive states or functions from behaviours. The restriction of coding to observable phenomena, and use of post hoc statistical analysis to identify links and associations between elements in the model, allows coding to be more objective and reliable. Retaining the narratives, as LOSA does, rather than the checklist and comments used in R-LOSA (Ma et al. 2011) allows for a fuller and richer contextual background to be preserved. This is essential for interpretation and coding of why actions made sense at the time for the observer, and for the verification team to ensure coding is consistent.

It is argued in this thesis that, despite these changes, NOM can still provide the claimed benefits of LOSA. Like LOSA, NOM can proactively investigate where risks may arise; highlight problems with procedures or working environments; provide data about which behaviours lead to safe outcomes; inform evidence-based safety interventions; and, ultimately, measure the effectiveness of interventions through ongoing monitoring.

Furthermore, NOM's simplified model and taxonomy is able to avoid many of the theoretical concerns about the original LOSA methodology. The result is a more robust data-collection tool that stays at the level of the observable, and a more representative model of human performance for collecting and organising data. When compared against the evaluation of LOSA for the same criteria, it can be argued that NOM offers improvements in areas such as comprehensiveness, consistency, acceptability and validity, as discussed in Chapter 11. On the basis of this evaluation, the NOM method met its objectives to provide novel insights for improving ground operations.

12.4 Implications for the NOM Model

Overall, the data broadly supports the assumptions put forward by the NOM model proposed in Chapter 7. The analysis of NOM data demonstrates that of all the possible associations between threats, variations and UOS that could occur, only some do, producing lawful relationships and insights into how incident opportunities develop. Although not explored further in this thesis, future analysis could investigate which variations are *less* likely where certain threats are present (shaded in pink in Table 24) to understand the implications of these negative correlations for the NOM model.

The results demonstrated a significant gap between how work is prescribed in rules and standard operating procedures and how it is performed in practice. Despite this variance, the ramp operations in this study operated within normal limits more than 87% of the time, which is in line with what the model might predict. Of particular interest, however, are the multiple pathways by which the operation can become unstable. About 12 per cent of observations fall into the abnormal category. The data confirms that some threats certainly do increase the likelihood of some variations, but also that threats are not necessarily required for errors to occur. Variations can and do occur in the absence of threats.

The data also supports the model's assumptions that variations increase the opportunity for Undesired Operational States to occur. This study also indicates, however, that variations are not necessary to produce undesired states. In this study, some undesired operational states result directly from threats (n22) where no management action was available to ramp crew. This discovery resulted in an amendment to the model to allow for threats to result in spontaneous undesired states.

The data from this study suggests that although mismanaged (18%) or unmanaged (92.1%) UOSs *can* lead to an accident, most do not, at least during the period of observation, with only one recordable accident being recorded during the turnaround in this study. All undesired operational states are, by definition, states that the organisation is seeking to avoid. It is not known whether the observed UOSs contributed to incidents outside of the turnaround, which is a possibility. For example, UOSs such as 'irregular loads' are also regulatory breaches which, if they travel, become recordable incidents at the port of discovery. As such, all UOSs are important to address. NOM provides a mechanism for identifying the pathways by which such states occur. The study identified some potential accident trajectories depicted by the model. The data provides a leading indicator to prevent accident trajectories through disrupting the associations between threats, variations and UOSs or through improving organisational defences.

The evaluation of the NOM model, compared to LOSA and TEM, suggests that it offers some improvements in critical areas. Overall, the data collected in this ramp implementation supported the concepts and relationships shown in the model in this environment. Future testing of NOM in other operational environments is recommended to understand its appropriateness for application in other industries.

12.5 Potential Application of NOM in Ground Safety and Beyond

Based on the discussion above, it can be concluded that NOM has been successful in meeting its development objectives. This research has met its aim of developing a data collection tool for ground safety that can provide a leading indicator of human and safety performance. These indicators provide novel, evidence-based insights to inform safety management systems and drive continuous improvement. NOM data could potentially lend itself to a number of specific applications now and in the future. The potential uses of NOM are discussed further below.

12.5.1 NOM data as a basis for industry learning and development

Cross-industry databases, such as the LOSA ARCHIE database, provide the potential for learning on a broader scale. Collecting data from many different sources requires a shared model and taxonomy of terms to allow data sharing. Dismukes (2010) suggests that TEM should be implemented through an industry-wide database to allow data to be entered in a similar format, and allow systematic exploration of issues cutting across the aviation industry. NOM, too, could be used for this purpose in the future. Chapter 2 discussed how the ground handling industry also struggles with data sharing at the industry level (Matthews 2012). NOM could be used as a framework for structuring industry-wide data collection from incidents or normal operations whilst protecting the data of individual organisations. This data could help inform industry-wide strategies and global targets for improvement, such as informing standards for airport design and layout or informing industry training standards.

12.5.2 NOM as an alternative to TEM/LOSA in aviation

This thesis argues that TEM and LOSA provided ground-breaking tools for collecting data about human and safety performance aviation, but also that the approach has been accepted and proliferated without sufficient scrutiny, particularly when applied outside the cockpit. When scrutinised in detail, TEM and LOSA were found to have some methodological weaknesses that may affect the reliability and usefulness of the data collected. NOM is an attempt to build on the strengths of LOSA whilst overcoming its methodological weaknesses. When compared with LOSA using the evaluation framework, NOM emerges as an enhanced methodology. It is therefore suggested that NOM be considered as an alternative to TEM and LOSA for aviation, including cockpit studies of normal operations.

12.5.3 NOM as a training tool

Concurrent with this study, the R-LOSA application discussed in Chapter 5 has informed the development of Ramp Resource Management training (RRM), which has now been launched as a proposed syllabus for ramp workers by the ground safety working group of the European Commercial Aviation Safety Team (ECAST). The training syllabus emphasises traditional CRM topics (human performance limitations, teamwork, leadership and communication etc.) but is tailored for ramp environment. The syllabus is publically available from EASA (Balk et al. 2012).

Future NOM monitoring could evaluate the effectiveness of this training by identifying whether the targeted behaviours have changed and the impact, if any, on desirable safety outcomes.

This research purposely did not attempt to measure a standardised range of CRM/NTS behaviours on the ramp, as there was insufficient evidence to support this approach. Rather, this study took the view that the observations would provide data to inform future training targets. The aim was to identify examples of mismanaged and managed threats and variations so they could become targets for behavioural improvement. Only a few management behaviours were identified in this study, but the data provided highlighted many examples where threats and errors could have been managed better. From here, it is possible to identify the skills and competencies that ramp crew might need to produce the desired management behaviours in the future, and to develop a tailored training program that addresses the issues identified. For example, based on this NOM implementation, industry training for ramp operations might focus on competencies such as managing communication difficulties, identifying and resolving discrepancies from load control, rejecting poorly presented freight and baggage, or improving threat and variation management through improved checks and supervision. Training programs that address airside driving risks and practices for all personnel who drive on the ramp would also be advantageous.

The data from this study also highlights where increased training would not be the most effective response, particularly where the issues highlighted were not the result of a lack of knowledge, competency or skill but related to system deficiencies. For example, problems with procedures, the design of the airport facilities, and operational pressures are best addressed at the system level. Understanding where and how training is best applied, and specifically what behaviours are needed to produce better outcomes based on evidence, could therefore be seen as an advantage offered by the NOM approach.

12.5.4 NOM as a self-report tool for normal operations or incidents

The TEM model has been used in many applications other than data collection, as discussed in Chapter 5. One application has been to organise and interpret data from incidents. Jones and Tesmer (1999) suggest that the TEM model is an effective framework for a human factors-centred incident reporting form and incident reporting questionnaire. TEM has also been adapted into a form of self-report for normal operations, such as the Unique Reporting Form (Leva et al. 2010). Flight crews provide a narrative or log of their flight, so that threats and hazards present during normal operations can be identified to provide quantitative data to inform the improvement of safety management processes (Leva et al. 2010, p. 165).

Like TEM, NOM could also be applied for this purpose, as a framework for reporting incidents or for structuring the collection of data from accident investigations. Such processes could help to consider how threats and variation management contributes to safety incidents. Operators could potentially be asked to provide narratives of their own normal operations, as well as identifying any threats or variations and undesired states. Future research could investigate how this data could identify where the safety management system needs to be strengthened.

12.5.5 Combining NOM data with other investigative approaches

The statistical analysis of associations between threats and variations in NOM offers some improvement to LOSA in determining the likelihood of two things occurring at the same time, and the narratives provide context and further insight as to why certain patterns may be occurring. These may provide a starting point for further analysis; however, further investigative methods may be required to understand cause and effect or to understand what is happening within a complex system. The advantage of NOM is that it provides new and invaluable data about where to look.

A recent example is provided by the United Airlines Data Visualisation Project, which uses LOSA data in combination with other data sources, including injury damage and near miss reports and audit data, to identify where work is not performed as planned and how this might be contributing to safety issues (Crocket 2017). At United Airlines, data from each of these databases was combined into a data warehouse. The information was statistically analysed for trends with a data visualisation overlay, using a program known as 'Spotfire'. The aim is to structure and represent data from multiple sources visually so that it can be easily interpreted according to stations or divisions. The vice president of Corporate Safety for United Airlines describes the objective of the project as to 'see if we can develop a program where we can take reams of data, map it and visualize it so that anybody from the CEO to the front line can look at it and say that's what our problems is' (Shay 2016). The aim in the future is make the program more predictive so that the system can predict threats that will occur based on variables such as bag loads, passenger load factors, gates out of service and weather information (Shay 2016).

In another example of LOSA data being combined with other investigative methods, De Boer et al. (2013) used a tool known as Systems-Theoretic Accident Model and Processes (STAMP) (Leveson 2004). STAMP was used to compliment R-LOSA data, to investigate the causes of the organisational strengths and weaknesses in a ground handling organisation.

De Boer (2017) reports that using the STAMP approach in platform safety, based on the work of Leveson (2011), has provided a number of insights in the gap between 'Work-as-Imagined and Work-as-Done' (Dekker and Hollnagel 2004). De Boer (2017) suggests that closing that gap requires management control actions based on accurate feedback on the ground handling process. De Boer suggests that both feedback and control actions are often lacking. Whilst programs like LOSA (and NOM) can contribute by giving insight into where rules are not complied with, STAMP goes a step further to help understand what is causing the gap.

STAMP has been applied in an airline ground service provider by Passenier et al. (2015) to investigate the causes of noncompliance of its work force, with some success. Although further investigation of STAMP is beyond the scope of the current research, the authors (Passenier et al. 2015, p. 36) suggest future research in this area should try to combine complementary methods such as R-LOSA with organisational ethnography and agent modelling. In order to improve the understanding of *why* work as performed is different to work as planned, NOM could also be combined with more investigative approaches, such as STAMP, ethnography or organisational systems modelling, to increase our understanding of the causes of human performance issues.

12.5.6 Using NOM data for HRA and risk assessment

Modern risk assessments require data about the likelihood of failure, including human failure, in order to estimate the risks associated with any high-hazard operation. Some high-hazard industries, such as nuclear chemical and rail industries, have developed domain-specific databases of human reliability to calculate these probabilities (Swain and Guttmann 1983; Boring and Guttmann 2010). With the exception of air traffic control (Shorrock and Kirwan 2002), however, there are few domain-specific methods for human reliability assessment in aviation (Harris et al. 2005). Although not developed specifically for this purpose, NOM data could be used to provide supporting data to these existing approaches. A NOM database, for example, could be used as a risk assessment tool based on real data, to produce error probabilities or create more accurate risk models. Although NOM in this study was adapted for the ground safety environment, the method could be applied as a way of collecting data about risk in any industry where behaviours and outcomes can be visually observed.

12.6 Summary and Conclusion

The ground handling industry faces a difficult challenge in improving its safety record during a period of intense growth in air traffic, the increase in composite aircraft, and ever-increasing operational and financial pressures. Reducing injury to personnel, damage to aircraft and ground-related delays are all priorities that need to be addressed urgently. NOM provides a tool that can potentially assist the industry in addressing these challenges. NOM can provide new information to monitor and measure safety in ground handling. It can identify relationships and associations in routine operations that allow accident trajectories to develop. The data provide a baseline measure of safety performance and identify high-value targets for intervention. Periodic NOM monitoring could also provide a measure of the effectiveness of interventions when compared to the baseline. NOM data can feed directly into an organisation's safety management system and offer a mechanism for measuring, and so driving, continuous improvement.

Furthermore, this study showed the benefits of a formal observational tool for identifying targets for action to improve safety in industry. NOM is offered as a generic tool that can be applied to collect data about normal operations in any observable task. Most high-hazard industries know that human performance is important to safety outcomes, but few routinely measure the standard of human performance in their operations. Many organisations only collect data about human performance issues by exception, through incident reports.

Normal Operations Monitoring can provide insights into events, situations and circumstances that have increased potential for risk, as well as the effectiveness or ineffectiveness of standard operating procedures. Beyond this, NOM can also provide a measure of system performance, to understand where the design of the environment or interfaces between service departments affect operational outcomes.

Although developed for a ground handling context, the NOM model and tools could potentially be applied in a range of high-hazard industries – in fact, any operation where performance is observable. The work in this thesis indicates that that NOM is an effective new tool for measuring human and safety performance in everyday ground operations, or indeed any high-hazard environment. Its application is likely to provide the data and insights necessary to drive continuous improvement to create safer operations, an aim for all industry.

Appendix 1: Categories of Ground Handling Codes for Threats, Variations and Undesired Operational States

1. Threat Codes

1.1 Airport/Facility

1.2 Airside Driving

1.3 Arriving Irregular Load

1.4 Bay Management

1.5 Communication

1.6 Documentation

1.7 Equipment

1.8 FOD

1.9 Freight/Baggage

1.10 Load Control

1.11 Operational Pressures

1.11 Weather

2 Variation Codes

2.1Bay Management Variations

2.2 Communication Variations

2.3. Documentation Variations

2.4 Door Variations

2.5 Driving Variations

2.6 Equipment Usage Variations

2.7 Loading Variations

2.7 Pushback Variations

2.8 Read back Variations

2.9 Restraining Variations

2.10 Unloading Variations

3. Undesired Operational State Codes

3.1 Conflict/Near Collision - Equipment

3.2 Conflict/Near Collision Equipment to Aircraft

3.3 Dangerous Goods

3.4 Delay

3.5 Ground Event

3.6 Injury/Near Miss to Personnel

3.7 Livestock Issues

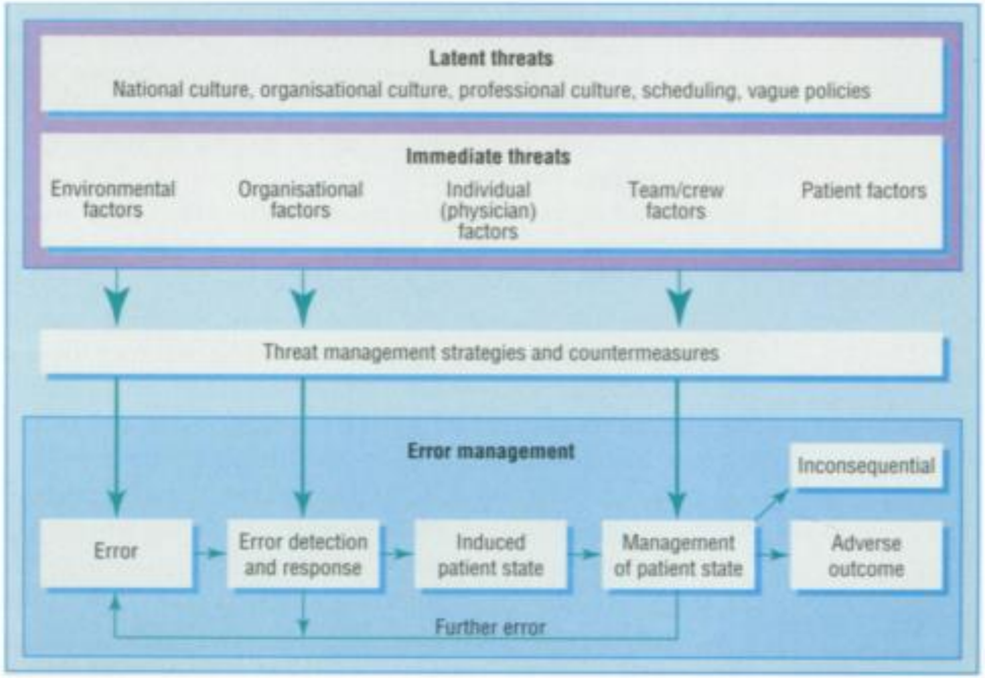
3.8 Loading

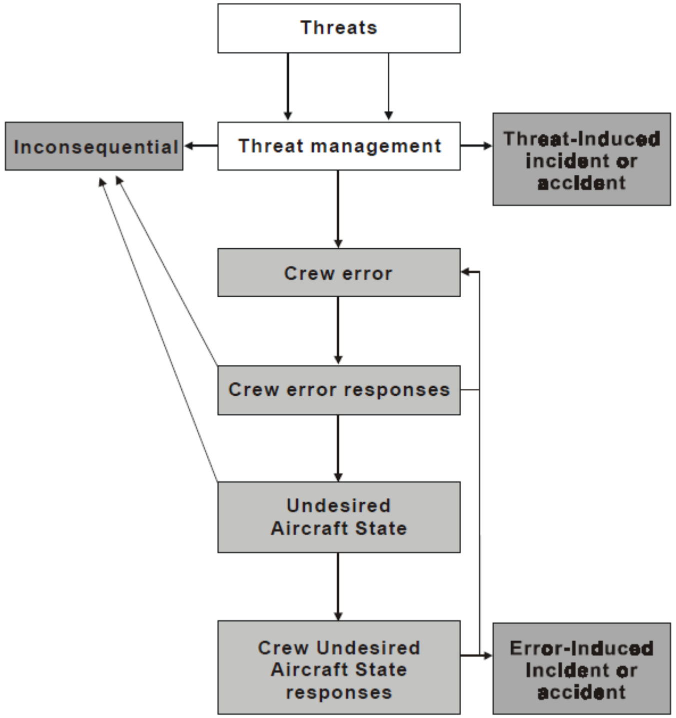
3.9 Pushback

Appendix 2 - LOSA Development Timeline

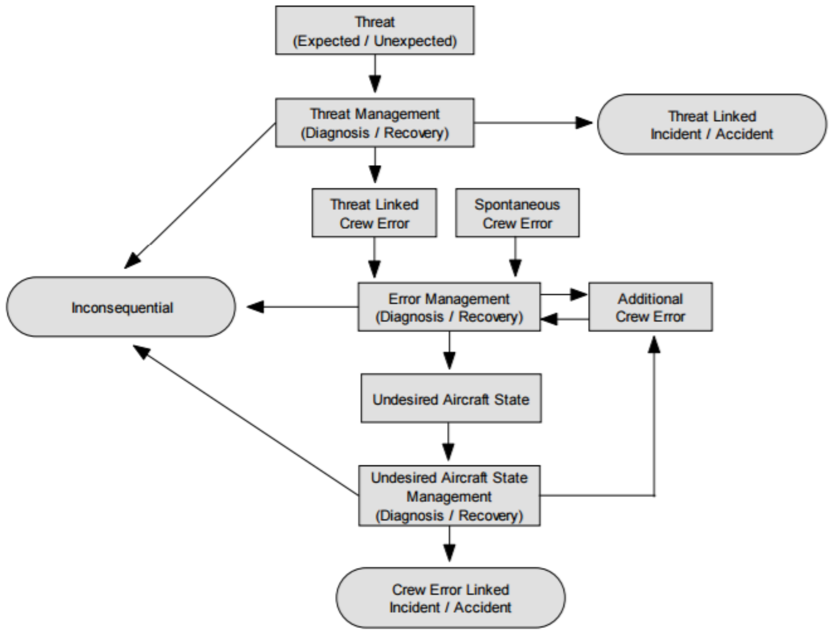
Date	Reference/Source	Description of TEM developments including: definitions of threats and errors, and description of the TEM model
1991	LOSA Collaborative Website: http://www.losacollaborative.org/history/	The LOSA research project began and was funded by the U.S. FAA.
1994	Merritt, Klinect, 2006 Defensive Flying for Pilots: An Introduction to Threat and Error Management	The precursor to LOSA began in 1994 at the request of Delta Air Lines. After developing a new Crew Resource Management (CRM) course for their line pilots, management questioned whether the concepts taught in training actually transferred to the line. A partnership between The University of Texas Human Factors Research Project (UT) and Delta Airlines developed a line audit methodology utilising jump-seat observations on regularly scheduled flights for a CRM audit methodology for normal operations.
1995	Klinect, J.R., Murray, P., Merritt, A. & Helmreich, R. (2003). Line Operations Safety Audit (LOSA): Definition and operating characteristics. In Proceedings of the 12th International Symposium on Aviation Psychology (pp. 663-668). Dayton, OH: The Ohio State University.	TWA, US Airways and American Airlines, and conducted their own CRM audits in collaboration with The University of Texas. The mid to late 1990's marked a paradigm shift - the project added threat and error management performance to its data collection... This shift in thinking fostered the development of the Threat and Error Management Model and the coining of the term Line Operations Safety Audit (LOSA) (Helmreich, Klinect Wilhelm & Merritt, 2001).
1 9 9 6	LOSA collaborative Website: http://www.losacollaborative.org/history/	First full TEM-based LOSA was conducted at Continental Airlines :Continental Airlines implements the first LOSA based on Threat and Error Management (TEM) and is recognized as first LOSA conducted by The LOSA Collaborative Members of The LOSA Collaborative would like to recognize the FAA for funding line operations safety audit and threat and error management research during our time at The University of Texas through AAR-100.

1999	Helmreich, R. L., Klinect, J. R., Wilhelm, J. A. (1999). Models of threat, error, and CRM in flight operations. In R.S. Jensen, B. Cox, J.F. Callister, R. Lavis (Eds.) *Proceedings of the Tenth International Symposium on Aviation Psychology* (pp. 298- 391). Columbus, OH, USA: Ohio State University.	CRM training became increasingly focused on threat and error management, where pilots were encouraged to identify, trap and manage errors before they could cause harm and the Threat and Error Management (TEM) model of human performance was introduced. The terminology in the TEM model, was further refined for use in pilot observations, 5 category taxonomy to classify the types of errors are outlined: ***Intentional non-compliance-*** conscious violations of Standard Operating Procedures (SOP's) or regulations. **Intentional noncompliance errors** are conscious violations of SOPs or regulations. Examples include omitting required briefings or checklists ***Procedural errors -*** when the intention is correct but the execution is flawed -describing situations in which procedures were followed but incorrectly executed. ***Communication errors-*** resulting from information being incorrectly transmitted or interpreted. ***Proficiency errors-*** - indicate a lack of knowledge or stick and rudder skill- (i.e. a pilot's lack of skill or knowledge in aircraft handling) ***Operational decision errors-*** discretionary decisions not covered by regulation and procedure that unnecessarily increases risk. Examples include extreme manoeuvres on approach, choosing to fly into adverse weather, or over-reliance on automation
1999	ICAO document 9803 2002, Appendix 2.2	ICAO endorses LOSA as the primary tool to develop countermeasures to human errors in aviation operations and makes the central focus its Flight Safety and Human Factors Program

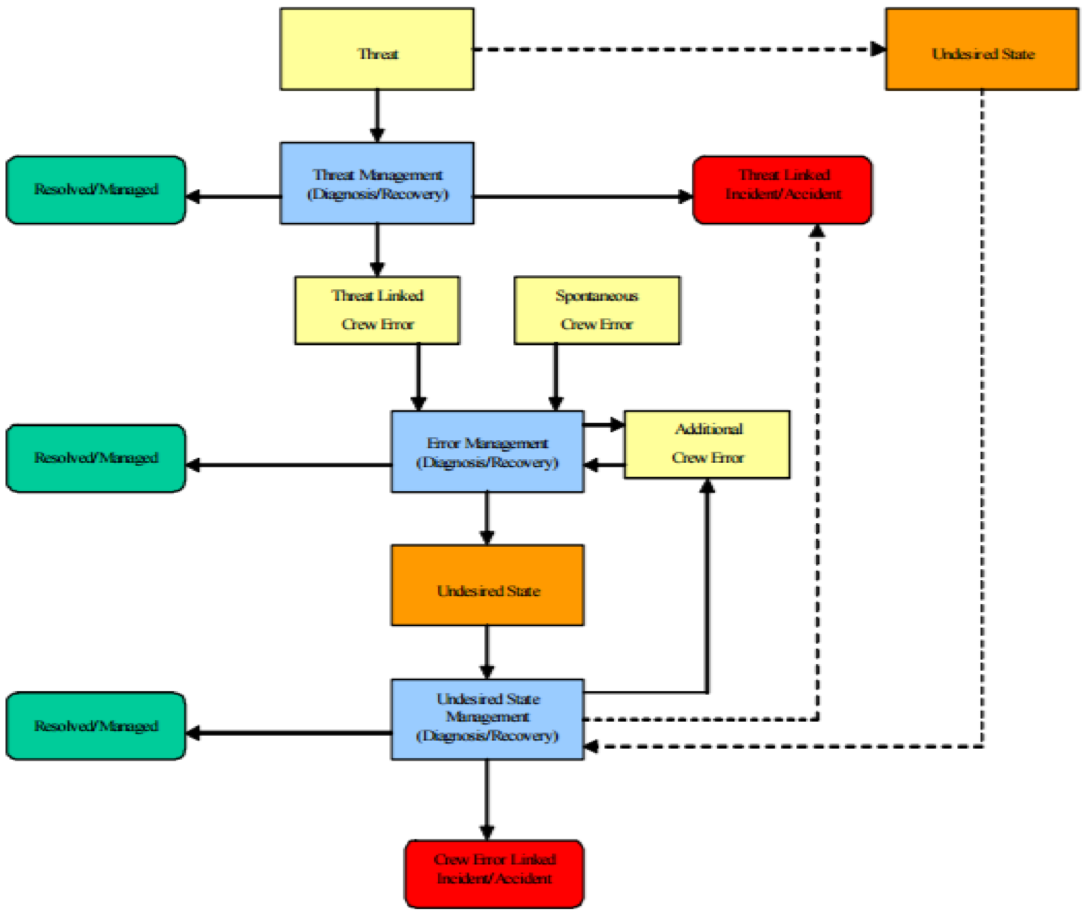
2000	Helmreich, R. (2000). On Error Management: Lessons from Aviation. *BMJ: British Medical Journal, 320*(7237), 781-785. Retrieved from http://www.jstor.org/stable/25187424	 **Fig 2** Threat and error model, University of Texas human factors research project
2001	Helmreich, R. L., Wilhelm, J. A., Klinect, J. R., & Merritt, A. C. (2001). Culture, error and crew resource management. In E. Salas, C. A. Bowers, & E. Edens (Eds.), *Improving teamwork in organizations:*	Helmreich et al. (2001) noted that after an error occurs there are three possible error management responses: *managed, mismanaged*, or *undetected/ignored*: ***Managed*** **s**ituations are those in which a threat or error is detected, trapped and dealt with successfully. ***Mismanaged*** situations occur when the threat or error is detected but the crew's response leads to a negative outcome which may then exacerbate the problem. ***Undetected/ignored*** situations occur when the crew fails to notice, react or respond to the error altogether. Threats and errors which are managed are described as ***inconsequential***, in that there are no adverse effects on the safe completion of the flight. However, outcomes are said to be ***consequential*** if the crew action results in further error, or undesirable aircraft states

	Applications of resource management (pp. 305-331). Hillsdale, NJ, USA: Erlbaum	
2002	ICAO Doc 9803 - International Civil Aviation Organisation. (2002). *Line operations safety audit – LOSA* (Doc. No. 9803 AN/761). Montreal, QE, Canada: International Civil Aviation Organisation.	International Civil Aviation Organization (ICAO) made LOSA a central focus of its Flight Safety and Human Factors Program and endorsed it as an industry best practice for normal operations monitoring (ICAO LOSA Manual, Doc 9803)ICAO recognizes LOSA as a Standard and Recommended Practice (SARP) with the publication of Line Operations Safety Audit guidelines which includes the following model of TEM in Appendix 2.2: 

		1**. Intentional non-compliance error**: Wilful deviation from regulations and/or operator procedures; 2. **Procedural error**: Deviation in the execution of regulations and/or operator procedures. The intention is correct but the execution is flawed. This category also includes errors where a crew forgot to do something; 3. **Communication error**: Miscommunication, misinterpretation, or failure to communicate pertinent information among the flight crew or between the flight crew and an external agent (for example, ATC or ground operations personnel); 4. **Proficiency error**: Lack of knowledge or psychomotor ('stick and rudder') skills; and 5. **Operational decision error**: Decision-making error that is not standardized by regulations or operator 2-4 Line Operations Safety Audit (LOSA) procedures and that unnecessarily compromises safety Links between threats and errors: Threats can link to errors and error can link to other errors or UAS as shown in the model. Errors that are caught in time do not produce negative consequences Ch 1 pages 1-5
2005	Annex 1 Personnel Licensing and Annex 6 Operations of Aircraft – http://www.losacollaborative.org/supporting-materials/	Threat and error management (TEM) is recognized and adopted by ICAO in Annex 1 Personnel Licensing and Annex 6 Operations of Aircraft
2006	Klinect, J. R. (2006). *Line operations safety audit: A cockpit observation methodology for monitoring commercial airline safety performance* (Doctoral dissertation). The University of Texas, Austin, USA	In his thesis, Klinect (2006) downplays the importance of quantifying the prevalence of threats, errors and UOS and stresses that it is the management of these phenomena that should be of particular interest (p. 78). Klinect also notes that less than 10% of cockpit crew errors had a threat identified as a precedent.

		 Page 16 /17 **Definitions of threats and errors:** **Threats –** Event error or aircraft state that occurs outside the control of flight crew but still requires their management to maintain safety **Error** – Crew Action or Inaction that leads to a deviation from crew or organisational intentions or expectations **Types of error**: **Aircraft handling**- directly linked to flying direction speed and configuration of the aircraft **Procedural Error** – pilot deviations from government regulations or standard airline operating procedures **Communication error** – poor or absent communication between pilots and or external agents There are subcategories of each presented on page 53 which include 5 different types of aircraft handling areas 7 types of procedural error and two categories of communication errors The thesis also provides codes within each category for '**intentional noncompliance**' although no definition could be found

		Possible responses : Detected and actioned – Crew detects error and actively tries to manage it Failing to respond – crew fails to detect or ignores an error leaving it unmanaged **Links between threats and errors:** In this model Klinect notes that 'errors can be spontaneous, linked to threats, or part of an error chain Errors can be inconsequential or link to another error or a UAS
2006	FAA Advisory Circular 120.90- Federal Aviation Administration. (2006). *Advisory Circular 120-90 – Line Operations Safety Audits*. Washington, DC, USA: United States Department of Transportation. Retrieved on Feb 13, 2015 from: http://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-90.pdf	The Federal Aviation Administration (FAA) also endorses LOSA as one of its voluntary safety programs (FAA Advisory Circular 12090 LOSA is formally recognized by the FAA as a Voluntary Safety program **Threat** - A threat is defined as an event or error that occurs outside the influence of the flight crew (i.e. it was not caused by the crew), increases the operational complexity of a flight, and requires crew attention and management if safety margins are to be maintained. **Error** - Crew error is defined as action or inaction that leads to a deviation from crew or organisational intentions or expectations. Errors in the operational context tend to reduce the margin of safety and increase the probability of adverse events. **Types of Error :** Aircraft handling errors procedural errors communication errors **Possible Responses** : Detected No Response Managed Mismanaged **Links between threats and errors:** No actual TEM model is shown however the document does reference The International Civil Aviation Organization (ICAO) TEM model in its Human Factors Training Manual (ICAO Document 9683, 2002). Appendix 3 (Error Management Worksheet) showing that only 1 of the 5 errors illustrated is linked to a threat, with the other 4 errors being spontaneous and in the absence of any reported threats
2006	Henry, C. (2007). The Normal Operations Safety Survey (NOSS): measuring system performance in air traffic control. Proceedings from t*he 2nd Institution of Engineering and*	Henry (2006) adapted a LOSA-like approach to gathering safety data in dispatch operations. **Threats** - External events or errors outside the influence of controllers, but which require their attention and management if safety margins are to be maintained **Errors** - Deviations from organisational expectations or controller intentions. (Page 25)

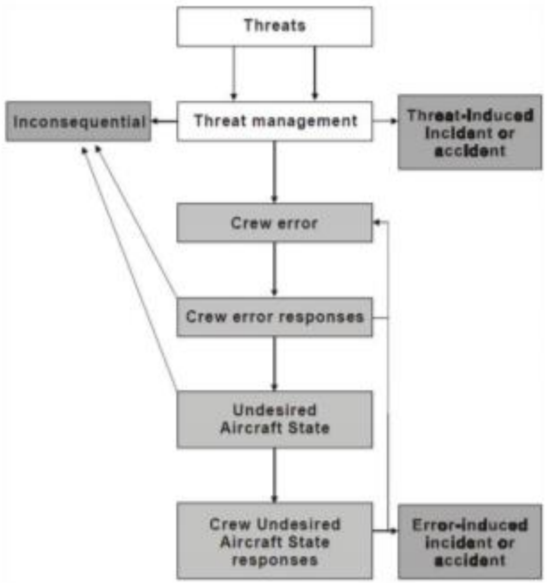
	Technology International Conference on Safety Management. Austin, TX, USA: University of Texas	 **Types of error:** Position Change Error, Communication,

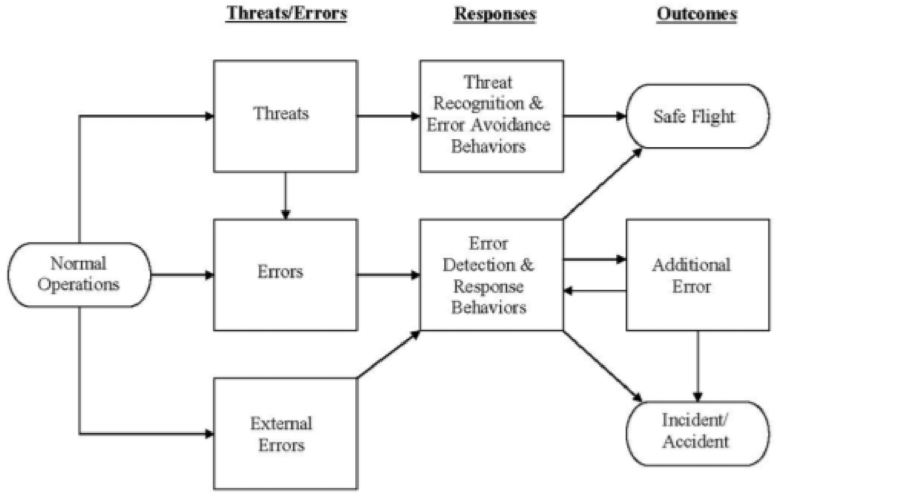
		Equipment Automation error, Flight data progress strips error, Procedural error, aircraft instruction error,
2006	Merritt, A., Klinect, J. (2006). *Defensive flying for pilots: an introduction to threat and error management*. The University of Texas Human Factors Research Project. Austin, TX, USA: The LOSA Collaborative.	30 commercial airlines in more than 14 countries around the world were applying the LOSA concept () **Threats** are defined as events or errors that: occur outside the influence of the flight crew (i.e. not caused by the crew); increase the operational complexity of a flight; and require crew attention and management if safety margins are to be maintained. **Error**: crew action or inaction that leads to a deviation from crew or organisational intentions or expectations types include: aircraft handling, procedural and communication errors. **Undesired Aircraft State**: a position, speed, attitude, or configuration of an aircraft that *results from flight crew error*, actions, or inaction; and clearly reduces safety margins. A safety-compromising state that results from ineffective error management.
2006	McDonald, A., Garrigan, B., & Kanse, L. C. (2006, February). Confidential observations of rail safety (CORS): An adaptation of line operations safety audit. Paper presented at the Proceedings of the Swinburne University Multimodal Symposium on Safety Management and Human Factors, Melbourne, Australia.	This adapted version of LOSA known as CORS refers to the original TEM model developed in Helmreich 1999 and outlined in ICAO 2002, p. 1: 'The Threat and Error Management Model (Helmreich et al., 1999) as used in aviation, defines threats as external situations that must be managed by the cockpit crew during normal, everyday flights. Threats increase the operational complexity of the flight and pose a safety risk to the flight at some level (ICAO, 2002). The operational definition of flight crew error on the other hand, is action or inaction that leads to deviation from crew and/or organisational intentions or expectations (Helmreich et al, 1999). Threats and errors are considered to be normal parts of everyday operations that must be managed. Threats and errors may sometimes go undetected, on other occasions they may be effectively managed, or may result in additional errors which require subsequent detection and response' LOSA's 10 Operating Characteristics were maintained for the study (page 2) and the authors describe how 'CORS observation materials follow a similar format to that of their LOSA equivalents however, the content of threat and error code lists is specific to the types of threats and errors that could be encountered in the rail domain. Threat and error code lists were generated using information gained through task analysis information, accident/incident investigation data, and focus groups with driver trainers and senior train crew. These lists were further developed during a trial phase of observations conducted as part of observer training. Threat and error lists were limited to those items that may impact on operational and/or passenger safety, and excluded those that would primarily impact on service efficiency or comfort. (page 2)
2008	CAAP 5.59-1(0) Teaching and Assessing Single-Pilot Human Factors and Threat and Error	The CASA CAAP reproduces the definition the University of Texas/GAPAN definitions and provides further definitions as follows: **Threat**: Events or errors that occur outside the influence of the flight crew; increase the operational complexity of the flight; and require crew attention and management if safety margins are to be maintained. **Threat (CASA modified definition for single pilot operations:** A situation or event that has the potential to impact negatively on the safety of a flight, or any influence that promotes opportunity for pilot error(s). This concept expands on the original definition

	Management Civil Aviation Advisory Publication October 2008	of threat and considers the psychological state of the pilot and the limitations they may bring with them to the aircraft on any given day. For example, increased levels of fatigue could result from a young child that is not sleeping well. **Error:** Flight crew actions or inactions that: lead to a deviation from crew or organisational intentions or expectations; reduce safety margins; and increase the probability of adverse operational events on the ground and during flight. **Types of Errors**: handling errors, procedural errors or communications errors. • **handling error** -when a pilot is interacting with an aircraft's controls, automation or systems. • **Procedural error** -when a pilot is using procedures such as checklists, SOPs or emergency actions. • **Communication error** -when pilots are interacting with other people such as ATC, ground assistants or other crew **Internal and External** - internal threats may lead to errors on the part of the pilot **Threat and Error Management (TEM**): The process of detecting and responding to threats and errors to ensure that the ensuing outcome is inconsequential, i.e. the outcome is not an error, further error or undesired state **Violation:** Intentional deviation from rules, regulations, operating procedures or standards. **Possible Responses:** **Inconsequential** means that there was no adverse outcome, i.e. there was not an error. **Consequential** may be a further error, or an undesired state. Other relevant resources include the CASA publications on Teaching and Assessing Single-Pilot Human Factors and Threat and Error Management (CAAP 5.59– 1(0)), Competency Based Training and Assessment in the Aviation Environment (CAAP 5.59a–1(0)), and Safety Behaviours: Human Factors for Pilots (see Reference section) These documents also refer to the ICAO 2002 doc as the original source document for LOSA/TEM
2011	Ma MJ, Pedigo, M., Blackwell, L., Hackworth, C., Holcomb, K., Gildea, K. (2011). *20 years of the line operations safety audit (LOSA) program: From flight operations to maintenance and ramp operations* (Doc. No. DOT/FAA/AM-11/15). Washington, DC, USA:	Airlines for America (A4A) Human Factors Task Force, which formed with the aim of developing tools and a method for ramp and maintenance LOSA. R-LOSA and M-LOSA are based on the TEM as described in the FAA 2006 publication. The resulting tools are known as R-LOSA and M-LOSA, respectively. The A4A Human Factors Task Force supported the development and testing of R- and M-LOSA at five U.S. airports in 2011, including a maintenance and ramp trial after which the task force released the R- and M-LOSA observation forms, procedures, databases and training materials.

	FAA, Office of Aerospace Medicine.	
2012	Ma MJ and Rankin W.L., (2012) Implementation Guideline for Maintenance Line Operations Safety Assessment (M-LOSA) and Ramp LOSA (R-LOSA) Programs Federal Aviation Administration, DOT/FAA/AM-12/9 Office of Aerospace Medicine Washington, DC 20591	 **Figure 1. Threat and Error Management Model** (Continental Airlines, 2008. Reprinted with permission.) The concepts of threats and errors are conceptually the same as in original cockpit LOSA, but adapted to the ramp working environment. **Errors** are defined as a 'mistake that is made when a threat is mismanaged' and also deviations from organisational expectations (FAA R-LOSA training P6) The training program include the LOSA error definitions Error types: Intentional non-compliance errors, procedural errors, communication errors, proficiency errors, and operational decision errors) (FAA R-LOSA training P6). **Links between threats and errors** 'The threat-error linkage is not necessarily straightforward, and it may not always be possible to establish a one-to-one mapping between threats and errors. Errors can be spontaneous without direct linkage to threats' (Ma and Rankin 2012 P1). In other aspects methodology appears to remain faithful to the original FAA LOSA /TEM approach as published in 2006 with some distinctions: Observers are asked not to speculate or guess what threats may have contributed to an error (p31) however observers are asked to link threats to their associated errors outcomes (i.e. Inconsequential, Undesired State, or Additional Error). Observers also code errors as being Inconsequential, leading to Undesired State, or Additional Error (P Ma and Rankin 2012 P40) **Checklists and Narratives** - The process of collecting data differs from the original methodology. Rather than complete narratives, R-LOSA and M-LOSA utilise structured observation checklists filled in by observers with supporting comments. The use of checklists is favoured for the standardisation of data and for ease of use and considered to be more effective than narratives in the ground domain (Ma et al 2011 P7). Results are entered into a ready-to-use database and tabulated by data-analysis software.

2012	De Boer, R. J., Koncak, B., Habekotte, R., Van Hilten, G. J. (2012). Introduction of ramp LOSA and KLM ground services, Amsterdam University of Applied Sciences, KLM Ground Services, Amsterdam, the Netherlands	De Boer et al. (2013) reported some difficulties in using the R LOSA tools published by the FAA. De Boer noted that the tools required considerable customisation to suit the KLM/Schiphol ramp environment. De Boer et al (2013) suggested, for example, that the R-LOSA suffered from a number of methodological problems in relation to the extensive checklist, which replaced the traditional narrative approach to LOSA observations From their own data, De Boer et al. (2013) suggested that errors may initiate with the ramp employees themselves and that the link between threats and errors may be more 'sporadic' than the original ICAO TEM model had indicated (p. 4). They concluded that, if this was the case, 'the identification and resolution of threats may only have a limited effect on safety' (p. 8). More recently De Boer (2017) suggested that the TEM theory has s number of flaws. De Boer raises concerns regarding the definition of a threat, suggesting that it lacks a clear unequivocal description in LOSA. De Boer also questions whether a threat is a really a condition which 'always needs management' and the fact that threats are identified 'after the fact' (i.e. with hindsight). De Boer (2017) also raises the point that LOSA and TEM assume a normative base for behaviour, and questions whether all variations from a norm should be considered undesirable. De Boer suggests instead that it may be more valuable to assess the gap between 'Work-as-Imagined and Work-as-Done' (De Boer 2017)
2013	Introduction of ramp-LOSA at KLM Ground Services Koncak, Habekotté, van Hilten2 (2013)	In 2013 De Boer et al suggested that the R-LOSA approach has methodological weaknesses 'The theory of threat and error management is as yet insufficiently mature to perform as a framework for ramp LOSA' (page 2). In particular, they criticised the 'questionable TEM framework' (p. 7). De Boer (2013) argues that the TEM model in the ICAO guidance material 'unjustly emphasises threats' (p7). The model is based on earlier models developed at the University of Texas at Austin Human Factors Project (Helmreich et al. 1999). The schematic representation of the model and the wording in the ramp LOSA guidance material imply that errors are a result of external or internal threats, such as faulty equipment, adverse weather (external) or fatigue (internal). However, 'errors can [also] be the result of a momentary slip or lapse' (Merritt & Klinect 2006) and therefore a threat need not necessarily precede an error. In fact, Klinect, Wilhelm et al. found for cockpit LOSA that less than 10% of crew errors had a threat as precedent (Klinect et al. 1999, p. 3) Criticisms include the description of threats as necessary precursors to errors and the suggestion that the identification of threats is susceptible to hindsight bias. The authors also suggested that the relationships reflected in the model may not reflect the reality of how threats and errors are experienced and managed in the ramp environment (p. 7). They suggest the TEM model may warrant further investigation and suggest consideration of more simplified models From their own data, De Boer et al. (2013) suggested that errors may initiate with the ramp employees themselves and that the link between threats and errors may be more 'sporadic' than the original ICAO TEM model had indicated (p. 4). They concluded that, if this was the case, 'the identification and resolution of threats may only have a limited effect on safety' (p. 8).

		 This study also refers to an unpublished model developed by Delta Airlines (2007) which shows how errors can be a consequence of threats and errors by others (external errors), but can also initiate from the ramp employees themselves.

		 Ramp LOSA definition: FAA provides a standard list of threat and error codes in their Ramp LOSA toolkit (FAA, 2010) Threat codes : generic threats in ramp environment **Error Codes** : defined as a deviation from a standard operating procedure,
2014	LOSA Archive Report: 10 Target Areas for Evidence Based Training IATA ITQI EBT Working Group Report 2014	LOSA Report for EBT Published by IATA refers to the TEM model adopted by ICAO and the FAA (Page 230) **Threat**: An event or error that occurs outside the influence of the crew but which requires the crew attention and management if safety margins are to be maintained. These are further sub divided in to environmental threats (outside the direct control of the airline) and airline threats (outside the direct control of the crew but within the purview of the airline management) **Flight Crew error** – An observed flight crew deviation from organisational expectations or crew intentions. **Error Types** include: Handling errors, (5 types) Procedural errors (7 types) Communication errors (2 types) Intentional non-compliance are defined as meeting one of 4 conditions such as an error that is committed multiple times within the same phase of flight, or the crew openly discusses the intention to violate, time optimising violations when time is available, or actions which increase risk when more conservative options are available (p. 245) **Possible responses:** Detected and acted upon Undetected/or not acted upon

		Detected with action No action taken This paper also notes that Error responses in LOSA are limited to what an observer can see in the cockpit without querying the flight crew. Those errors not acted upon are assumed to be ignored or undetected.
2016	The LOSA Collaborative Website http://www .losacollaborative.org	By 2016, the LOSA collaborative lists over 45 airlines on their website who have applied LOSA and further endorsements from the Flight Safety Foundation, International Air Transport Association (IATA) and The International Federation of Air Line Pilots Associations (IFALPA)

References

Airbus 2013 – *Global Market Forecast- Future Journeys* Report published by AIRBUS S.A.S. 31707 Blagnac Cedex, France

Airport Cooperative Research Program. (2011). *Synthesis of ramp safety practices*. Sponsored by the Federal Aviation Administration. Washington, DC, USA: Transportation Research Board.

Alexander, M., Cox, S., Cheyne, A. (1994). Safety culture within a UK based organisation engaged in offshore hydrocarbon processing. Proceedings from *The Fourth Conference on Safety and Well-being at Work.* Loughborough, UK: Centre for Hazard and Risk Management, pp. 64-75.

Amalberti, R. (1996). *La conduite des systèmes à risques*. In Hoc, J. M. *Le Travail Humain*. Paris: Presses Universitaires de France.

Amalberti, R. (2001). The paradoxes of almost totally safe transport systems. *Safety Science, 37*(2-3), 109-126.

Ananda, CM, Kumar, R, Ghoshhajra M (2010) *Configurable flight data analysis for trends and statistics analysis-an embedded perspective of an efficient flight safety system,* EEE International Conference on Aerospace Electronics, Communication and Instrumentation 6-7 Jan 2010 Vijayawada

Angell, I. O., Straub, B. (1999). Rain-dancing with pseudoscience. *Cognition, Technology and Work, 1*, 179-196.

Apostolakis, G. E. (2004). How useful Is quantitative risk assessment? *Risk Analysis, 24*(3), 515–520.

Applegate, J. D., & Graeber, R. C. (2005). Integrated safety system design and human factors considerations for jet transport aeroplanes. In D. Harris and H. C. Muir (Eds.), *Contemporary issues in human factors and aviation safety* (pp. 3-23). Aldershot, UK: Ashgate Publishing.

Australian Transport Safety Bureau. (2010). *Aircraft loading occurrences July 2003 to June 2010* (Publication No. ATSB-Dec10/ATSB156. Investigation No. AR-2010-044). Canberra, ACT, Australia: Australian Government.

Australian Transport Safety Bureau. (2010). *Evaluation of the human factors analysis and classification system as a predictive model* (Investigation No. AR-2008-036). Canberra, ACT, Australia: Australian Govt. (Authors: Inglis, M., Smithson, M. J., Cheng, K., Stanton, D. R., Godley, S. T.).

Australian Transport Safety Bureau. (2010). *Ground operations occurrences at Australian airports 1998 to 2008* (Publication & Investigation No. AR-2009-042). Canberra, ACT, Australia: Australian Government.

Australian Transport Safety Bureau. (2011). *A systematic review of the effectiveness of safety management systems – ATSB transport safety report* (Doc. No. XR-2011-002, author Thomas, M. J. W.). Canberra, ACT, Australia: Australian Government.

Aven, T., Renn, O. (2009). On risk defined as an event where the outcome is uncertain. *Journal of Risk Research Volume, 12*(1), 1–11.

Balk, A. (2008). *Safety of ground handling.* National Aerospace Laboratory The Netherlands:

Balk, A. D., Boland, E. J., Nabben, A. C., Bossenbroek, J. W., Clifton-Welker, N., Landauer, N., Rolfsen, J., Schaffner, B., Bellamy, L. J., Geyer, T. A. W., Wilkinson, J. (2008). *Ramp resource management training syllabus development.* Amsterdam, The Netherlands: National Aerospace Laboratory & the ECAST Ground Safety Working Group.

Balk, A. D., Bossenbroek, J. W. (2010). *Aircraft ground handling and human factors: A comparative study of the perceptions by ramp staff and management*. Amsterdam, The Netherlands: National Aerospace Laboratory.

Barach, P. Small, S. D. (2000) Reporting and preventing medical mishaps: Lessons from non-medical near miss reporting systems, *British Medical Journal, International edition* 320.7237 (Mar 18, 2000): 759-63.

Bao, S., Howard, N., Spielholz, P., Polissar, N. (2009). Interrater reliability of posture observations. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 51*(33), 292-309.

Bartlett, J. K., Kotrlik, J. W., Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal, 19*(1), 43.

Bass, M. (1995) Surfaced detection system for airports - US5469371 Report Ser. No. 07/811,817, filed Dec. 20, 1991, now U.S. Pat. No. 5,375,058. University Of Central Florida

Bedford, T., Cooke, R. (2001). *Probabilistic risk analysis: Foundations and methods*. Cambridge, UK: Cambridge University Press.

Bellamy, L. J., Beyer, T. A. W., Wilkenson, J. (2008) Development of a functional model which integrates human factors, safety management systems and wider organisational issues. *Safety Science, 46*(3), 461-492.

Bird, F. E., & Loftus, R. G. (1976). *Loss control management*. Loganville, GA, USA: Institute Press.

Bisantz, A. M., & Drury, C. G. (2005). Applications of archival and observational data. In J. R. Wilson & N. Corlett (Eds.), *Evaluation of human work* (3rd ed.). Boca Raton, FL, USA: CRC Press

Bisantz, A. M., Ockerman, J. J. (2002). Informing the evaluation and design of technology in intentional work environments through a focus on artefacts and implicit theories. *International Journal of Human-Computer Studies, 56*(2), 247-265.

Bishop R , Murray , D, McDonald W, Hickman J, Bergoffen G. - transportation research board 2008 "Safety Impacts of Speed Limiter Device Installations on Commercial Trucks and

Buses " *Federal Motor Carrier Safety Administration Repor*t ISSN 1544-6808 ISBN: 978-0-309-09827-4 Library of Congress Control Number 20089112791

Blackman, H. S., Gertman D. I., Boring R. L. (2008). Human error quantification using performance shaping factors in the SPAR-H method. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52*(21), 1733–1737.

Boeing. (2000a). Maintenance error decision aid (MEDA) users guide. Chicago, IL, USA: Shappell, S. A. & Wiegmann, D. A.

Boeing. (2000b). Ramp error decision aid (REDA) users guide. Chicago, IL, USA: Shappell, S. A. & Wiegmann, D. A.

Borys, D. (2000). Seeing the wood from the trees: a systems approach to OHS management. Proceedings from *Occupational Health and Safety Management Systems: the First National Conference*. University of Western Sydney, NSW, Australia.

Bottani, E., Monica, L., Vignali, G. (2009). Safety management systems: Performance differences between adopters and non-adopters. *Safety Science, 47*(2), 155-162.

Bourrier, M. (2002). Bridging research and practice: the challenge of normal operations studies. *Journal of Contingencies and Crisis Management*, *10*(4), page 173-180.

Bradburn, N. M., Sudman, S. (1980). *Improving interview method and questionnaire design: Response effects to threatening questions in survey research*. San Francisco, CA, USA: Jossey-Bass.

Brand, F. S., Jax, K. (2007). Focusing the meaning(s) of resilience: Resilience as a descriptive concept and a boundary object. *Ecology and Society, 12*(1), 23.

Braithwaite, Wears, Hollnagel, (2016) "*Reconciling Work-as-Imagined and Work-as-Done* ", Resilient Health Care, Volume 3: Reconciling Work-as-Imagined and Work-as-Done August 25, 2016 by CRC Press, Reference - ISBN 9781498780568 - CAT# K29926

Brown, O. & Hendrick, H.W. (Eds.). (1996). *Human Factors in Organizational Design and Management – V*. Amsterdam, The Netherlands: Elsevier. (pp. 667-672).

Brown, K.A. Willis, PG. , Prussia G.A. (200) Predicting safe employee behavior in the steel industry: development and test of a sociotechnical model *Journal of Operations Management,* 18 (2000), pp. 445–465

Burdekin, S. G. (2003) Mission Operations Safety Audits (MOSA): Measurements of behavioural performance or non-technical skills from military aircrew. *Aviation Safety* spotlight 0403.

Burdekin S. G. (2014). 'MOSA Comes of Age: the transition of an experimental research project to validate aircrew behavioural self-reports into a mature operational tool', Proceedings from *11th International Symposium of the Australian Aviation Psychology Association*. Melbourne, Australia.

Cacciabue, P. C. (2005). Human error risk management methodology for safety audit of a large railway organisation. In J. Wilson, B. Norris, S. Clarke, & A. Mills (Eds.), *Rail human factors: Supporting the integrated railway*. Cornwall, UK: Ashgate Publishing.

Capa Centre for Aviation2014 “Part 1: Liberalisation, efficiency & compensation” *Airport ground handling – industry overview*. 18-Nov-2014

Carder, B., Ragan, P. (2003). A survey-based system for safety measurement and improvement. *Journal of Safety Research, 34*(2), 157–165.

Carletta, J., Anderson, A. H., McEwan, R. (2000). The effects of multimedia communication technology on non-collocated teams: a case study. *Ergonomics, 43*(8), 1237–1251.

Catchpole K & Jeffcott S “*Human Factors and Ergonomics Practice in Healthcare*” Chapter 13 – in Human Factors and Ergonomics in Practice: Improving System Performance and improving performance and wellbeing in the real world Steven T. Shorrock editor New York Routledge, 2016

Chandler, J. G. (1995). Putting a dent in ground damage. *Aviation Equipment Maintenance 14*(6), 22-25.

Chassin M and Loeb J (2013) High – Reliability Health Care: Getting there from here, *The Milbank Quarterly*, Volume 91, Issue 3, pages 459–490, September 2013

Ciavarelli, A. Jr., Figlock, R. (1996). Organizational factors in aviation accidents. Proceedings from *The Ninth International Symposium on Aviation Psychology* (pp. 1033-1035). Columbus, OH, USA.

Ciavarelli, A. P. (1997). *Safety climate and risk culture: How does your organization measure up?*. Monterey, CA, USA: Naval Postgraduate School. Retrieved on 13 Oct 2014 from: http://calhoun.nps.edu/bitstream/handle/10945/35044/Ciavarelli_OSES_FSF_safety%20 climate-culture-REPRINT_with.pdf?sequence=1

Ciavarelli, A. P. (2005). *Assessing safety climate and culture: From aviation to medicine*. St. Louis, MS, USA: St. Louis University.

Ciavarelli, A. P., Figlock, R. (1997). Organizational factors in naval aviation accidents. In R. S. Jensen & L. A. Rakovan (Eds.) *Proceedings of the ninth international aviation psychologists conference* (pp. 1033-1039). Columbus, OH, USA: Ohio State University.

Cinnamon, J., Schuurman, N., Hameed, S. (2011, Jun). Pedestrian injury and human behaviour: Observing road-rule violations at high-incident intersections. *PLoS One, 6*(6).

Civil Aviation Safety Authority (CASA). (1988). CAO 82.5 CAAP SMS - Civil Aviation Order 82.5 - Conditions on air operators' certificates authorising regular public transport operations in high capacity aircraft (12/12/2004). *Civil Aviation Act 1988* (Paragraph 28BA (1) (b)and subsection 98 (4A)). Canberra, ACT, Australia: Legislative Drafting Branch, Legal Services Division, Civil Aviation Safety Authority.

Civil Aviation Safety Authority. (2004 August 23). Safety Oversight Program (ASOP). *Aviation Week and Space Technology, 161*(8), 78.

Civil Aviation Safety Authority (October 2008) CAAP 5.59-1(0*) Teaching and Assessing Single-Pilot Human Factors and Threat and Error Management* Civil Aviation Advisory Publication; October 2008

Civil Aviation Safety Authority. (2009). *Safety Management Systems for Regular Public Transport Operations*. Canberra, ACT, Australia: Civil Aviation Advisory.

Civil Aviation Safety Authority. (2010). *Civil Aviation Order 82.5 Amendment Order (No.2)*. Canberra, ACT, Australia: The Australian Government. Retrieved from: http://www.casa.gov.au/wcmswr/_assets/main/lib100017/cao82_5-2.pdf

Civil Aviation Authority (2011) CAAP SMS-3(1) Non-Technical Skills Training and Assessment for Regular Public Transport Operations Civil Aviation Advisory Publication April 2011

Clark, M. C., Czaja, S. J., Weber, R. A. (1990). Older adults and daily living profiles. *Human Factors, 32*, 537-549.

Clarke, S. (1999). Perceptions of organizational safety: implications for the development of safety culture. *Journal of Organizational Behaviour, 20*, 185-198.

Clarke, S. (2000). Safety culture: under-specified and overrated?. *International Journal of Management Reviews, 2*(1), 65–90.

Clarke, S. (2003). Safety climate in an automobile manufacturing plant: the effects of work environment, job communication and safety attitudes on accidents and unsafe behaviour in an automobile manufacturing plant. *Personal Review, 35*(4), 413-430.

Coleman, N. (2009). *A380 Introduction into service; Human performance in the aircraft turnaround*. Internal presentation to Qantas Airports Services. Mascot, Sydney, Australia: Human Ergonomics Australia.

Connell, I. W. (1998). Error analysis of ticket vending machines: comparing analytic and empirical data. *Ergonomics, 41*(7), 927-961.

Connolly, T. J., Backwell, B. B. (1987). A simulator approach to training in aeronautical decision-making. In R. S. Jensen (Ed.) *Proceedings of the fourth international symposium on aviation psychology* (pp. 251-258). Columbus, OH, USA: Ohio State University.

Cook, R. I., Woods, D. D. (1996). Adapting to new technology in the operating room. *Human Factors, 38*(4), 593-613.

Cooper, D. F. (2004). The Australian and New Zealand Standard on Risk Management, AS/NZS 4360: 2004.' Tutorial Notes: Broadleaf Capital International Pty Ltd (2004).

Cooper, D., Finley, L. McClure, J. & Frend, D. (2013). *Documented business benefits from implementing a behaviour-based safety process*. BSMS Inc., Franklin, IN, US. Retrieved on 12 Apr 2015 from: http://www.behavioral-safety.com/articles/Documented_Business_Benefits_of_BBS_Processes.pdf

Cooper, G. E., White, M. D., & Lauber, J. K. (Eds.). (1980). *Resource management on the flightdeck: Proceedings of a NASA/Industry workshop (NASA CP-2120)*. Moffett Field, CA, USA: NASA-Ames Research Centre.

Cooper, M. D. (2000). Towards a model of safety culture. *Safety Science, 36*(2), 111–136.

Cooper, M. D. (2002, Jun). Safety culture: A model for understanding and quantifying a difficult concept. *Professional Safety, 47*(6), 30-36.

Cooper, M. D., Phillips, R. A. (2004). Exploratory analysis of the safety climate and safety behaviour relationship. *Journal of Safety Research, 35*(5), 497–512.

Cox, S. J., Flin, R. (1998). Safety culture: philosopher's stone or man of straw? *Work and Stress, 12*(3), 189-201.

Craig, H. (November, 2006). Cathay Pacific LOSA. Paper presented at the *Fourth ICAO-IATA LOSA & TEM Conference*. Toulouse, France: ICAO-IATA.

Cranswick, L. (2010). *Extracts from H.W. Heinrich, "Industrial Accident Prevention", 3rd edition, 1950*. Retrieved on 3 Jun 2015 from: http://lachlan.bluehaze.com.au/books/nsc_safety_facts_1941/

Crocket L (2016) Personal communication regarding United Airlines use of R LOSA data email received 11 January 2017 courtesy of Lisa Crocket, Sr. Manager, Quality Assurance, LOSA United Corporate Support Centre

Croft, J. (2001, July 16). Researchers perfect new ways to monitor pilot performance. *Aviation Week & Space Technology*, pp. 76-77.

Dalijono, T. (2006). Reducing human error by improvement of design and organization. *Process Safety and Environmental Protection, 84*(B3), 191–199.

De Boer R.J., B. Koncak, R. Habekotté, and G.-J. van Hilten,(2011) "Introduction of ramp-LOSA at KLM Ground Services," in Human Factors of Systems and Technology, 2011.

De Boer, R. J., Koncak, B., Habekotte, R., Van Hilten, G. J. (2012). Introduction of ramp LOSA and KLM ground services, Amsteram Univerity of Applied Sciences, KLM Ground Services, Amsterdam, the Netherlands

De Boer, R. J., de Jong, S. (2013). Application of STAMP to facilitate interventions to improve platform safety. In *Proceedings of the 3rd STAMP Workshop*, Boston, NY, USA.

De Boer R J, (2017) Personal Communication via email 24, January 2017

De Carvalho, P. V. R. (2011). The use of Functional Resonance Analysis Method (FRAM) in a mid-air collision to understand some characteristics of the air traffic management system resilience. *Reliability Engineering & System Safety, 96*(11), 1482–1498.

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

Dekker, S. W. A. (2001). The disembodiment of data in human factors accident analysis. *Journal of Human Factors and Aerospace Safety, 1*(1), 39–57.

Dekker, S. W. A. (2002). *The field guide to human error investigations*. Hampshire, UK: Ashgate Publishing Ltd.

Dekker, S. W. A. (2003). Illusions of explanation: A critical essay on error classification. *The International Journal of Aviation Psychology, 13*(2), 95-106.

Dekker S.W.A., and Hollnagel E (2004) "Human factors and folk models." *Cognition, Technology & Work* 6.2 (2004): 79-86.

Dekker, S. W. A. (2007). Why doctors are more dangerous than gun owners: A rejoinder to error counting. *Human Factors, 49*(2), 177–184.

Deming, W. Edwards (1986). *Out of the Crisis*. Cambridge, MA, USA: MIT Press. ISBN 0-911379-01-0

Dijkstra A (2007) Resilience Engineering and Safety Management Systems in *Aviation*, January 2007

Dillingham, G. L. (2007). *Aviation Runway and Ramp Safety: Sustained Efforts to Address Leadership, Technology, and Other Challenges Needed to Reduce Accidents and Incidents*. Washington, DC, USA: U.S. Government Accountability Office.

Dillman, D. A., Sinclair, M. D., Clark, J. R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly, 57*(3), 289-304.

Dismukes, R. K. (2009). *Human error in aviation*. Aldershot, UK: Ashgate Publishing Ltd.

Dismukes, R. K. (2010). Understanding and analyzing human error in real-world operations. In E. Salas and D. Maurino (Eds.), *Human factors in aviation* (2nd ed.) (Chapter 11, pp. 335-374). Burlington, MA, USA: Elsevier.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives, 38*(1), 105-123. Retrieved from: www.ccsenet.org/journal/index.php/jsd/article/download/31886/19993

Drury, C. G. (1995). The use of archival data. In J. R. Wilson and E. N. Corlett (Eds.) *Evaluation of human work* (2nd ed.) (Chapter 5, pp. 101-112). New York, NY, USA: Taylor and Francis.

Drury, C. G., Wenner, C. A., Murthy, M. R. (1998). *Phase VIII progress report – 1.2.4. Evaluation of ground damage interventions*. Buffalo, NY, USA: State University of New York

Duncan, N. K. (2013). *A delphi study of decision support systems for aviation safety management program cost estimation* (Doctoral dissertation #3569897). Northcentral University, AZ, USA. Retrieved from: http://gradworks.umi.com/35/69/3569897.html

Earl, L., Murray, P. S., Bates, P. R. (2011). Line Operations Safety Audit (LOSA) for the management of safety in single pilot operations in Australia and New Zealand. *Aeronautica, 1*(1). 1-7.

Edwards, E., (1972) Man and machine: systems for safety. In: *Proceedings of British Airline Pilots Associations Technical Symposium*. British Airline Pilots Associations, London, pp. 21–36.

Ek, Å., Akselsson, R. (2005). Safety culture on board six Swedish passenger ships. *Maritime Policy & Management, 32*(2), 159–176.

Ek, Å., Akselsson, R. (2007). Aviation on the ground: safety culture in a ground handling company. *The International Journal of Aviation Psychology, 17*(1), 59–76.

Ek, Å., Akselsson, R., Arvidsson, M., & Johansson, C. R. (2002). Safety culture in the Swedish air navigation services. In C. Weikert, E. Torkelsson & J. Pryce (Eds.), *Proceedings of the Fourth European Conference of the European Academy of Occupational Health Psychology* (pp. 58–61). Dec 4-6, Vienna, Austria.

Embrey, D. (1986). SHERPA: a systematic human error reduction and prediction approach. In *International Topical Meeting on Advances in Human Factors in Nuclear Power System* (pp. 184–193). Knoxville, TN, USA: American Nuclear Society.

Embrey, D. E. (1992). Incorporating management and organisational factors into probabilistic safety assessment. *Reliability Engineering and System Safety, 38*. 199-208.

Embrey, D. E. (1993). Quantitative and qualitative prediction of human error in safety assessments. *Institute of Chemical Engineers Symposium Series, 130*, 329–350.

Enterprise Risk Management (ERM). (2010). *A structured approach to Enterprise Risk Management (ERM) and the requirements of ISO 31000*. Devon, UK: Alarm – The Public Risk Management Association, AIRMIC & IRM.

EUROCONTROL (2009) "A White Paper on Resilience Engineering for ATM" *European Organisation for the Safety of Air Navigation* (EUROCONTROL). September 2009

Federal Aviation Administration(FAA). (2006-2013). *Safety Management Manual AC 120-92 4*. Washington, DC, USA: US Government Printing Office.

Fahlbruch, B., & Wilpert, B. (1999). System safety—An emerging field for I/O psychology.

Federal Aviation Administration. (1988). *Advisory Circular: System Design and Analysis (AC 25.1309-1A)*. Washington, DC, USA: United States Department of Transportation.

Federal Aviation Administration (FAA). (2006). *Advisory Circular 120-90 – Line Operations Safety Audits*. Washington, DC, USA: United States Department of Transportation. Retrieved on Feb 13, 2015 from: http://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-90.pdf

Federal Aviation Administration (FAA). (2007). *Exploration of the correlation structure of a survey for evaluating airline safety culture*. Washington, DC, USA: Gibbons, A. M., von Thaden, T. L., Wiegmann, D. A.

Federal Aviation Administration (FAA). (2012). *Implementation guideline for Maintenance line operations safety assessment (M-LOSA) and ramp line operations safety assessment (R-LOSA) programs* (Doc. No. DOT/FAA/AM-12/9). Washington, DC, USA: Ma, M., Rankin, W. L.

Federal Aviation Administration (FAA). (2012, March 8). *Press release – Airline passenger travel to nearly double in two decades*. Price, H. J. Retrieved from: https://www.faa.gov/news/press_releases/news_story.cfm?newsId=13394

Ferrel R 1997 as cited in 2.3.2.1 section on "Ferrel Theory" page 57 in Hosseinian, Seyyed Shahab, and Zahra Jabbarani Torghabeh. "Major theories of construction accident causation models: a literature review." *International Journal of Advances in Engineering & Technology* 4.2 (2012): 53-66.

Fernandez-Muniz, B., Montes-Peon, J. M., Vazquez-Ordas C. J. (2009). Core elements of safety culture and safety performance: Literature review and exploratory results. *International Journal of Society Systems Science, 1*(3), 227-259,

Fernández-Muñiz, B., Montes-Peón, J. M., Vázquez-Ordás, C. J. (2007). Safety culture: analysis of the causal relationships between its key dimensions. *Journal of Safety Research, 38*(6), 627-641.

Feyer, A. M., Williamson, A. M. (1991). A classification system for causes of occupational accidents for use in preventive strategies. *Scandinavian Journal of Work, Environment & Health, 17*(5), pp. 302-311.

Feyer, A.M, Williamson, A.M. (1998). Rules or Trust; Ensuring Compliance. In A. Feyer & A. Williamson (Eds.) *Occupational Injury, Risk, Prevention and Intervention* (pp. 215-217). London, UK: Taylor & Francis.

Fleming, M. (2000). Safety culture maturity model. In Health and Safety Executive's *Offshore Technology Report*. London, UK: The Keil Centre for the Health and Safety Executive.

Fleming, M. T., Flin, R. H., Mearns, K., Gordon, R. P. E. (1996). The offshore supervisor's role in safety management: Law enforcer or risk manager. Proceedings from *the 3rd International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production* (SPE paper 35906). New Orleans, LA, USA.

Fletcher G., Flin, R. H., 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*(5). 580-588.

Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N., Patey, R. (2004). Rating non-technical skills: Developing a behavioural marking system for use in anaesthesia. *Cognition, Technology & Work, 6*(3), 165-171.

Flight Safety Foundation (FSF). (2004) The ground accident prevention project. *Aviation Week and Space Technology, 161*(8), 78 August 24 2004

Flight Safety Foundation (FSF). (2007). *Ground Accident Prevention Data* (online calculator). Retrieved on Dec 4 2014 from: http://flightsafety.org/archives-and-resources/ground-accident-prevention-gap

Flight Safety Foundation (FSF) (2007) "Diffusing the Ramp- Progress report on efforts to stem the toll of Ground Accidents, Ground Safety Report – Lacagnina *Airosafety world magazine* May 2007

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*(2), 95-117.

Flin, R., Mearns, K., O'Connor, P., Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science, 34*, 177-192.

Frantzi, J. (2009, May-June). Aero-illustrations reprinted in M. Marchbank (Ed.) Flight safety starts on the ground. *Flight Safety Australia, 68*, 8-15 (p.10).

Freeman, C., Simmon, D. A. (1991). Taxonomy of crew resource management: Information processing domain. In R. S. Jensen (Ed.), *Proceedings of the Sixth Annual International Symposium on Aviation Psychology* (pp. 391 – 397). Columbus, OH, USA: Ohio State University.

Frey, K. S., Hirschstein, M., Endstrom, L., Snell, J. (2009). Observed reductions in school bullying, non-bullying aggression, and destructive bystander behavior: A longitudinal evaluation. *Journal of Educational Psychology, 101*(2), 466.

Frick, P. J., Christopher, T. B., Kamphaus, R. W. (2010). Structured diagnostic interviews. In P. J. Frick, T. B. Christopher, R. W. Kamphaus (Eds.) *Clinical assessment of child and adolescent personality and behavior* (pp. 253-270). Manhattan, NY, USA: Springer US.

Frigg, R., Hartmann, S. (2006). Models in science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy online* (2012). Retrieved from: http://plato.stanford.edu/archives/fall2012/entries/models-science/)

Frigg, Roman, Hartmann, Stephan. (2006). Models in science. In Edward N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2006 Edition). Retrieved April 26 2014 at: http://plato.stanford.edu/entries/models-science/

Fuller, R., McDonald, N., White, G., Walsh, W. (1994). Strategies to improve human performance safety in ground operations. Proceedings from *The Airports Council International Apron Safety Seminar*. Caracas, Venezuela.

Gaudart, C. (2000). Conditions for maintaining aging operators at work—a case study conducted at an automobile manufacturing plant. *Applied Ergonomics, 31*(5), 453-462.

Geertz, C. (1973). Toward an interpretive theory of culture. In C. Geertz (Ed.), *The Interpretation of Cultures: Selected Essays* (pp. 3-30). New York, NY, USA: Basic Books Inc.

Goritz, A. (2004). The impact of material incentives on response quantity, response quality, sample composition, survey outcome and cost in online access panels. *International Journal of Market Research, 46*(3), 327-346.

Graeber, R. C. (2006, November). The safety benefits of LOSA: A manufacturer's view. Paper presented at *The Fourth ICAO-IATA LOSA & TEM Conference.* Toulouse, France: ICAO-IATA.

Grote, G., 2009. Response to Andrew Hopkins. *Safety Science* 47 (2009) 478

Guldenmund, F. W. (2000). The nature of safety culture: a review of theory and research. *Safety Science, 34,* 215-257.

Gunther, D. (2002). Threat and error management training counters complacency, fosters operational excellence. *International Civil Aviation Organisation Journal, 57*(4), 12–13.

Hackworth, C., Holcomb, K., Gildea, K., Blackwell, L., Ma, M., Pedigo, M. (2010). Proactive safety management: Maintenance and Ramp LOSA. Presentation to *Regional Airline Association Conference* May 24-27. Washington, DC, USA: FAA.

Haddon W. (1980). Options for the prevention of motor vehicle crash injury. *Israel Journal of Medical Sciences, 16*(1), 45-68.

Haddon, W. (1970). On the escape of tigers: an ecologic note. *American Journal of Public Health, 60*(12), 2229-2234. DOI: 10.2105/AJPH.60.12.2229-b.

Haddon, W. (1973). Energy damage and the ten countermeasure strategies. *Journal of Trauma, 13*(4), 321-331. DOI: 10.1097/00005373-197304000-00011.

Hale, A. R. (2003). Safety management in production. *Human Factors and Ergonomics in Manufacturing & Service Industries, 13*(3), 185–201.

Hale, A. R., Guldenmund, F. W., van Loenhout, P. L. C. H., Oh, J. I. H. (2010). Evaluating safety management and culture interventions to improve safety: Effective intervention strategies. *Safety Science, 48*(8), 1026-1035.

Hale, A. R., Heijer, T. (2006). Is resilience really necessary? – The case of railways. In D. D. Woods, N. Leveson, E. Hollnagel (Eds.) Resilience engineering: Concepts and precepts. Aldershot, UK: Ashgate Publishing Ltd.

Hale, A. R., Heming, B. H. J., Carthey, J., Kirwan, B. (1997). Modelling of safety management systems. *Safety Science, 26*(1), 121-140.

Harris, D., Stanton, N. A., Marshall, A., Young, M. S., Demagalski, J., Salmon, P., (2005). Using SHERPA to predict design-induced error on the flight deck. *Aerospace Science and Technology, 9*(6), 525-532. DOI: 10.1016/j.ast.2005.04.002

Hawkins, F. H. (1987). *Human factors in flight* (2nd ed.). Aldershot, UK: Ashgate Publishing.

Health And Safety Executive (1999). *The Implementation of CORE-DATA, a computerised human error probability database*. HSE Contract Research Report 245/199. Sheffield, UK: Gibson, W. H., Megaw, E. D.

Health and Safety Executive. (2001). *Reducing Risks, Protecting People: HSE's decision-making process.* Norwich, UK: HMSO.

Health and Safety Executive. (2013). *A study of HSE RIDDOR and enforcement statistics for the air transport industry - 2009/10 to 2013/14*. Retrieved from: http://www.hse.gov.uk/airtransport/sic2007.pdf

Health and Safety Executive. (2013). *Managing for Health and Safety* (Series code: HSG65). London, UK: HSE publications.

Heese, M. (2012). Got the results, now what do you do? Safety culture transformation from theory to practice. *Aviation Psychology and Applied Human Factors, 2*(1), 25-33.

Heinrich, H. W. (1929). The foundation of a major injury. *National safety News, 19* (1), 9–11, 59.

Heinrich, H. W. (1931). *Industrial accident prevention: A scientific approach*. New York, NY, USA: McGraw-Hill.

Heinrich, H. W., Petersen, D., & Roos, N. (1980). *Industrial accident prevention: A safety management approach* (5th ed.). New York, NY, USA: McGraw-Hill.

Helmreich, R. (2000). On error management: lessons from aviation. *British Medical Journal, 2000*(320), 745-749.

Helmreich, R. L. (2001). Assessing system safety. In *Safety in Aviation: The Management Commitment: Proceedings of a Conference*. London, UK: Royal Aeronautical Society.

Helmreich R. L. (2002). Crew performance monitoring programme continues to evolve as database grows. *International Civil Aviation Organisation Journal, 57*(4), 6-7.

Helmreich, R. L. (2006). Beyond the cockpit: The spread of LOSA and threat and error management. Paper presented at *The Fourth ICAO-IATA LOSA & TEM Conference*. Toulouse, France.

Helmreich, R. L., & Merritt, A.C. (2000). Safety and error management: The role of crew resource management. In B. J. Hayward & A. R. Lowe (Eds.) *Aviation resource management* (pp. 107-119). Aldershot, UK: Ashgate Publishing.

Helmreich, R. L., Klinect, J. R., Wilhelm, J. A. (1999). Models of threat, error, and CRM in flight operations. In R.S. Jensen, B. Cox, J.F. Callister, R. Lavis (Eds.) *Proceedings of the Tenth International Symposium on Aviation Psychology* (pp. 298- 391). Columbus, OH, USA: Ohio State University.

Helmreich, R. L., Wilhelm, J. A., Klinect, J. R., & Merritt, A. C. (2001). Culture, error and crew resource management. In E. Salas, C. A. Bowers, & E. Edens (Eds.), *Improving teamwork in organizations: Applications of resource management* (pp. 305-331). Hillsdale, NJ, USA: Erlbaum.

Henley, E. J, & Kumamoto, H. (1981). *Reliability engineering and risk assessment*. Englewood Cliffs, NJ, USA: Prentice-Hall. 1985

Henley, E. J, & Kumamoto, H. (1985). *Designing for reliability and safety control*. Englewood Cliffs, NJ, USA: Prentice-Hall.

Henry, C. (2007). The Normal Operations Safety Survey (NOSS): measuring system performance in air traffic control. Proceedings from t*he 2nd Institution of Engineering and Technology International Conference on Safety Management*. Austin, TX, USA: University of Texas.

Hoffman, R., Crandell, B., Shadbolt, N. (1988). Use of CDM to elicit expert knowledge a case study in the methodology of cognitive task analysis. *Human Decision Processes 62*(2), 129-158.

Hofstede, G. (1982). *Values survey module*. Maastricht, The Netherlands: Institute for Research on Intercultural Cooperation.

Hofstede, G. (1983). *Culture and management development*. Geneva, Switzerland: Geneva International Labour Office.

Hofstede, G. (1984). *Culture's consequences; international differences in work-related values* (abridged ed.). London, UK: Sage Publications.

Hofstede, G. (1994). *VSM 94: Values Survey Module 1994*. Retrieved September 17, 2014, from http://stuwww.uvt.nl/~csmeets/~1st-VSM.html

Hofstede, G. (1997). *Culture and organisations: Software of the mind*. NY, USA: McGraw Hill.

Hollnagel, E. (1998). *Cognitive reliability and error analysis method – CREAM*. Oxford, UK: Elsevier Science.

Hollnagel, E. (1998). Cognitive reliability and error analysis method: CREAM (1st ed.). Oxford, NY, USA: Elsevier.

Hollnagel, E. (2009). *Safer complex industrial environments: A human factors approach*. Boca Raton, FL: CRC Press.

Hollnagel, E. (2009). *The ETTO principle: Efficiency-thoroughness trade-off*. Farnhamn, UK: Ashgate Publishing Ltd.

Hollnagel E. (2011) *Understanding Accidents, or How (Not) to Learn from the Past*- Presented by University of Southern Denmark available from: http://www.functionalresonance.com/FRAM-1_understanding_accidents.pdf

Hollnagel, E. (2012). A tale of two safeties. *Nuclear Safety and Simulation, 4*(1), 1-9. Retrieved from: http://www.resilienthealthcare.net/A_tale_of_two_safeties.pdf

Hollnagel, E. (2015). *RAG – Resilience Analysis Grid – Introduction to the Resilience Analysis Grid (RAG)*. Retrieved on July 26 2015 at: http://erikhollnagel.com/onewebmedia/RAG%20Outline%20V2.pdf

Hollnagel, E., & Amalberti, R. (Eds.). (2001). The emperor's new clothes, or whatever happened to "human error". Proceedings from *The 4th International Workshop on Human Error, Safety and Systems Development*. Linköping, Sweden.

Hollnagel, E., & Embrey, D. E. (1994). *Human reliability assessment in safety assessments addressing: Cognitive and action errors*. Parbold, UK: HRA Ltd.

Hollnagel, E., & Woods, D. D. (2006). Epilogue: resilience engineering precepts. In Hollnagel, E., Woods, D. D., Leveson, N., *Resilience engineering: Concepts and precepts* (pp. 3-7). Aldershot, UK: Ashgate Publishing.

Hollnagel, E., Nemeth, C.P. & Dekker, S. (Eds.). (2008). *Resilience engineering perspectives* (Vol. 1, remaining sensitive to the possibility of failure). Aldershot: Ashgate Publishing.

Hollnagel, E., Pedersen, O. M., Rasmussen, J. (1981). *Notes on human performance analysis*. Published by Roskilde, Denmark: Riso National Laboratory,

Hollnagel, E., Woods, D. D. (2005). Joint cognitive systems: foundations of cognitive systems engineering. Boca Raton, FL, USA: CRC Press/Taylor & Francis.

Hollnagel, E., Woods, D. D., & Leveson, N. (2007). *Resilience engineering – Concepts and precepts*. Aldershot, UK: Ashgate Publishing.

Hopkins, A. (2000). *Lessons from Longford: the Esso gas plant explosion*. North Ryde, NSW, Australia: CCH Australia Ltd.

Hopkins, A. (2007). Thinking About Process Safety Indicators, Working Paper 53. Paper presented at *The Oil and Gas Industry Conference*. Manchester, UK: National Research Centre for OHS regulation.

Hovden, J., Albrechtsen, E., Herrera, I. A. (2008). Is there a need for new theories, models and approaches to occupational accident prevention? *Safety Science, 48*(8), 950–956.

Hoveden, J., Tinmannsvik, R. K. (1990). Internal control: a strategy for occupational safety and health: experiences from Norway. *Journal of Occupational Accidents, 12*(1-3), 21-30.

Howard, S. K., Gaba, D. M., Fish, K. J, Yang, G., Sarnquist, F. H. (1992). Anaesthesia crisis resource management training: Teaching anaesthesiologists to handle critical incidents. *Aviat Space Environ Med, 63*(9), 763-770.

Howe, J. (2001). *Behavior-based safety can be hazardous to your health and safety program* (version 3). UAW Health and Safety Department. Retrieved from: http://www.uawlocal974.org/BSSafety/Warning!_Behavior-Based_Safety_Can_Be_Hazardous_To_Your_Health_and_Safety_Program!.pdf

Howell, J., Miller, P., Park, H. H., Sattler, D., Schack, T., Spery, E., Widhalm, S., & Palmquist, M. (2012). Reliability and Validity. In *Writing@CSU Writing Guide*. Colorado State University. Retrieved from: http://writing.colostate.edu/guides/guide.cfm?guideid=66

Hsiao, Y.L., Drury, C., Wu, C., Paquet, V. (2013). Predictive models of safety based on audit findings: part 2: Measurement of model validity. *Applied Ergonomics, 44*(4), 659–666.

Hudson, P. (2003). Applying the lessons of high risk industries to health care. *Quality and safety in health care*, *12*(suppl 1), i7-i12.

International Air Transport Association (IATA) (2015) Annual Review 2015 presented at *71st IATA Annual General Meeting* Miami, June 2015

International Air Transport Association (IATA), ISAGO Standards Manual GOSM Ed 5, March 2016, ISBN 978-92-9252-793-8 Montreal—Geneva

International Atomic Energy Agency. (1998). *Collection and classification of human reliability data for use in probabilistic safety assessments: Final report of a co-ordinated research programme 1995-1998*. (Ref. No.: IAEA-TECDOC-1048). Vienna, Austria.

International Civil Aviation Authority. (1989). *Human Factors Digest No. 1 Fundamental Human Factors Concepts* (Circular number 216-AN/131). West Sussex, UK: International Civil Aviation Organisation.

International Civil Aviation Organisation (ICAO). (2013). *Safety Management Manual (SMM) 12.1.4 (Doc No. 9859)*. ICAO.

International Civil Aviation Organisation. (1983). Part 8, Airport operational services. In *Airport services manual* (Doc. No. 9137 AN/898). Montreal, QE, Canada: International Civil Aviation Organisation.

International Civil Aviation Organisation. (2001). *Manual on certification of aerodromes* (Doc. No. 9774 AN/969). Montreal, QE, Canada: International Civil Aviation Organisation.

International Civil Aviation Organization. (2002a). Line Operations Safety Audit (LOSA). ICAO (Document 9803 AN/761). Montreal, QE, Canada: System Safety : International Civil Aviation Organisation

International Civil Aviation Organisation. (2004). *International Standards and Recommended Practices (4th ed.). Annex 14 (Aerodromes), Volume 1*. Montreal, QE, Canada:

International Civil Aviation Organisation. (2006). *Safety management manual – SMM* (Doc 9859 AN/460). Montreal, QE, Canada: International Civil Aviation Organisation.

International Civil Aviation Organisation. (2009). *Safety management manual – SMM* (2nd ed.) (Doc. No. 9859 AN/474). Montreal, QE, Canada: International Civil Aviation Organisation.

International Civil Aviation Organisation. (2010). *Human error in aviation maintenance the years to come*. Maurino, D. E. Retrieved from: http://www.plant-maintenance.com/articles/human_error_aviation_maintenance.pdf

International Civil Aviation Organization. (2013). *Annex 19 to the convention on International Civil Aviation: Safety management*. Montreal, QE, Canada: ICAO.

Isaac, A., Shorrock, S., Kirwan, B. (2002). Human error in European air traffic management: the HERA project. *Reliability Engineering and System Safety, 75*(2), 257-272.

ISAGO (2015). *ISAGO Standards Manual 1st – 5th edition (2010- 2015)*. Montreal, QE, Canada — Geneva, Switzerland: International Air Transport Association (IATA).

ISAGO. (2015). *IATA Safety Audit for Ground Operations because flight safety starts on the ground*. Faisal-ejaz. Retrieved on Oct 2, 2015, from: http://docslide.us/documents/isago-info-brochure-oct-2010.html.

ISO/IEC. (2003). ISO Standard 12100 – *Safety of machinery, basic concepts, general principles for design. Part one: Basic terminology and methodology*. Geneva, Switzerland: International Organisation for Standardisation (ISO).

Janssen, W. (1994). Seat belt wearing and driving behaviour: An instrumented-vehicle study. *Accident Analysis and Prevention, 26*(2), 249–261.

Jepsen, R. M. H. G., Østergaard, D., Dieckmann, P. (2015). Development of instruments for assessment of individuals' and teams' non-technical skills in healthcare: a critical review. *Cognition Technological Work, 17*(1), 63–77.

Jiao, M., Wood, B. (2012). Maintenance and ramp line operations safety assessment (M&R LOSA): Using a tested approach for risk management and safety enhancement. *Aviation in Focus, 3*(1), 44-52.

Jones, S. G., Tesmer, B. (1999). A new tool for investigating and tracking human factors issues in incidents. In R.S. Jensen, B. Cox, J.F. Callister, R. Lavis (Eds.) *Proceedings of the Tenth International Symposium on Aviation Psychology*. Columbus, OH, USA: Ohio State University.

Kearns, S. K., Sutton, J. E. (2013). Hangar talk survey using stories as a naturalistic method of informing threat and error management training. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55*(2), 267-277.

Kennedy, R., Kirwan, B. (1995). The failure mechanisms of safety culture. In A. Camino, G. Weimann (Eds.), *Proceedings of the international topical meeting on safety culture in nuclear installations* (pp. 281-290). Vienna, Austria: American Nuclear Society of Austria.

Kerkloh, M. I. (1992). The socio-psychological consequences of personal injuries on the ramp – the Frankfurt experience. Paper presented at the *International Air Transport Association Seminar -- Creating the Safety Culture on the Ramp*, Miami, FL, USA: Federal Aviation Administration.

Kirwan, B. (1992a). Human error identification in human reliability assessment – part 2: detailed comparison of techniques. *Applied Ergonomics, 29*(5), 299–318.

Kirwan, B. (1992b). Human error identification in human reliability assessment: part 2 – detailed comparison of techniques. *Applied Ergonomics, 23*(6), 371–381.

Kirwan, B. (1994). *A guide to practical human reliability assessment*. London, UK: Taylor & Francis.

Kirwan, B. (1996). Human error identification techniques for high risk systems – part 1: review and evaluation of techniques. Applied Ergonomics, 29(3), 157-177.

Kirwan, B. (1996). The validation of three human reliability quantification techniques – THERP, HEART and JHEDI: Part I – Technique descriptions and validation issues. *Applied Ergonomics, 27*(6), 359–374.

Kirwan, B. (1997a). Validation of human reliability assessment techniques: part 1 – validation issues. *Safety Science, 27*(1), 25-41.

Kirwan, B. (1997b). Validation of human reliability assessment techniques: part 2 – validation results. *Safety Science, 27*(1), 43-75.

Kirwan, B. (1998). Human error identification techniques for risk assessment of high risk systems: part 1 – review and evaluation of techniques. *Applied Ergonomics, 29*(3), 157–177.

Kirwan, B. (1999). The development of TRACEr: A technique for the retrospective analysis of cognitive errors in ATM. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics (Vol. 3)*. Hants, UK: Ashgate Publishing.

Kjellén, U. (1984). The deviation concept in occupational accident control – part I: definition and classification. *Accident Analysis & Prevention, 16*(4), 289–306.

Kjellén, U. (1995). Integrating analyses of the risk of accidents into the design process – part II: method for prediction of the LTI rate. *Safety Science, 19*, 3-18.

Kjellén, U. (2000). *Prevention of accidents through experience feedback*. London, UK: Taylor & Francis.

Kjellén, U. (2009). The safety measurement problem revisited. *Safety Science, 47*(4), 486–489.

Kletz, T. (1974). *HAZOP and HAZAN – Notes on the identification and assessment of hazards*. Rugby, UK: Institute of Chemical Engineers.

Klinect J. R. (2002). LOSA searches for operational weakness while highlighting systemic strengths. *International Civil Aviation Organisation Journal, 57*(4), 8-9.

Klinect, J. R. (2006). *Line operations safety audit: A cockpit observation methodology for monitoring commercial airline safety performance* (Doctoral dissertation). The University of Texas, Austin, USA.

Klinect, J. R., Murray, P., Merritt, A., & Helmreich, R. (2003). Line operations safety audit (LOSA): definition and operating characteristics. In *Proceedings of the 12th international symposium on aviation psychology* (pp. 663-668). Dayton, OH: Ohio State University.

Klinect, J. R., Wilhelm, J. A., Helmreich, R. L. (1999). Threat and error management: data from line operations safety audits. In R.S. Jensen, B. Cox, J.F. Callister, R. Lavis (Eds.) *Proceedings of* *the Tenth International Symposium on Aviation Psychology* (pp. 683-688). Columbus, OH, USA: The Ohio State University.

La Porte, T. R. (1996). High reliability organizations: Unlikely, demanding and at risk. *Journal of Contingencies and Crisis Management, 4*(2), 60-71.

La Porte, 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 – 47.

Lacagnina, M. (2007, May). Defusing the ramp. *Aerosafety World, 11*, 20-24.

Landau, M. (1969). Redundancy, rationality and the problem of duplication and overlap. *Public Administration Review*, *29*(4), 346–358.

Landry, J., Ingolia S. (2011). *ACRP Synthesis 29: Ramp safety practices*. Washington, DC, USA: Transportation Research Board of the National Academy of Science.

Laughery, K., Kent, P. V. (1993). Major and minor injuries at work: Are the circumstances similar or different? *International Journal of Industrial Ergonomics, 12*(4), 273–279.

Lay, E., Matthieu, B. (2014). Noticing brittleness designing for resilience. In C. P. Nemeth, E. Hollnagel in *Engineering in practice, volume 2: Becoming resilient*, (chapter 10 Page 138 - 140). Aldershot, UK: Ashgate Publishing Ltd.

Le Coze, J. C., Pettersen, K. (2008). Is resilience engineering realist or constructivist? Proceedings from Symposium ’08: *3rd Resilience Engineering International Symposium*. Antibes, France: Resilience Engineering Association. Retrieved from: http://www.resilience-engineering-association.org/download/resources/symposium/symposium-2006/Le_Coze_Pettersen_text.pdf

Learmount, D. (2005). Ground control: Why is an industry that is so safety conscious in the air so lax when it comes to accidents on the ramp? Flight International, July 2005

Leva, M. C., Cahill, J., Kay, A. M., Losa, G., McDonald, N. (2010). The advancement of a new human factors report - 'The Unique Report' - facilitating flight crew auditing of performance/operations as part of an airline's safety management system. *Ergonomics, 53*(2), 164–183.

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science, 42*(4), 237-270.

Li, W. C., Harris, D. (2006). Pilot error and its relationship with higher organizational levels: HFACS analysis of 523 accidents. *Aviation, Space and Environmental Medicine, 77*(10), 1056-1061.

Li, W. C., Harris, D., Yu, C. S. (2008). Routes to failure: analysis of 41 civil aviation accidents from the Republic of China using the human factors analysis and classification system. *Accident Analysis & Prevention, 40*(2), 426-434.

Liao, Meng-Yuan. "Safety Culture in commercial aviation: Differences in perspective between Chinese and Western pilots." *Safety science* 79 (2015): 193-205.

Liou, J.H., Tzeng, G.H., Chang, H.C., 2007. Airline safety measurement using a hybrid model. *Journal of Air Transport Management* 13, 243–249

Liou J, Yenb, L Tzeng, GH. (2008) Building an effective safety management system for airlines, *Journal of Air Transport Management*, Volume 14, Issue 1, January 2008, Pages 20–26

Ljungberg, D., Lundh, V. (2013). *Resilience engineering within ATM - Development, adaption, and application of the Resilience Analysis Grid (RAG).* Norrköping, Sweden: Department of Science and Technology, Linköping University.

Longstaff, P. H., Koslowskib, T. G., Geogheganc, W. (2013). Translating resilience: A Framework to Enhance Communication and Implementation

López de Castroa, B, a, F J, a, JM, Pietrantonic L., Hernándeza, A. (2013) Testing the validity of the International Atomic Energy Agency (IAEA) safety culture model, *Accident Analysis & Prevention* , Volume 60, November 2013, Pages 231–244

LOSA Archive Report (2014) 10 Target Areas for Evidence Based Training IATA ITQI EBT Working Group Report 2014 (available from: http://www.losacollaborative.org/supporting-materials/)

Loukopoulos, L. D., Barshi, I. (2003). Opportunities for and vulnerabilities to error in everyday flight operations. Presented at *Air Transport Association/Crew Resource Management Conference,* Bloomington, MN, USA.

Ma M.J (2016) Personal communication received 21 December 2016, courtesy of Maggie Ma PhD, CHFP, Maintenance Human Factors, Commercial Aviation Services at The Boeing Company based regarding information shared at the FAA Safety Infoshare sessions occurring between 2014 -2015)

Ma, MJ., Pedigo, M., Blackwell, L., Gildea, K., Holcomb, K., Hackworth, C. (2011). *The line operations safety audit program: Transitioning from flight operations to maintenance and ramp operations*. (Publication No. DOT/FAA/AM-11/15). Oklahoma City, OK, USA: Federal Aviation Administration, Civil Aerospace Medical Inst. Retrieved from: http://www.faa.gov/library/reports/medical/oamtechreports/2010s/media/201115.pdf

Ma, M.J., Pedigo, M., Blackwell, L., Hackworth, C., Holcomb, K., Gildea, K. (2011). *20 years of the line operations safety audit (LOSA) program: From flight operations to maintenance and ramp operations* (Doc. No. DOT/FAA/AM-11/15). Washington, DC, USA: Federal Aviation Administration, Office of Aerospace Medicine.

Ma, M.J., and Rankin W.L., (2012) Implementation *Guideline for Maintenance Line Operations Safety Assessment (M-LOSA) and Ramp LOSA (R-LOSA) Programs* Federal Aviation Administration, DOT/FAA/AM-12/9 Office of Aerospace Medicine Washington, DC 20591

Ma, M.J,Wood, B., (Jun/Jul, 2012). Maintenance and ramp line operations safety assessment (M&R LOSA): using a tested approach for risk management and safety enhancement. *Aviation in Focus, 3*(1), 44-52.

Mahajan, R. P. (2010). Critical incident reporting and learning. *British Journal of Anaesthesia, 105*(1), 69-75.

Manuele, F. A. (2002). *Heinrich revisited: Truisms or myths*. Itasca, NY, USA: National Safety Council Press.

Marchbank, M. (Ed.). (2009, May-June). Flight safety starts on the ground. *Flight Safety Australia, 68,* 8-15.

Marchbank, M. (2011). Civil Aviation Safety Authority of Australia. (2011, Nov – Dec). Safety starts on the ground. *Flight Safety Australia Magazine, 83*, 9-15.

Marx D (2017)- *There is No Such Thing as a High Reliability Organization,* What we believe Volume 1 Issue 1 , Published on February 1, 2017 available form : https://www.linkedin.com/pulse/thing-high-reliability-organization-david-marx?trk=prof-post&trk=hp-feed-article-title-share&utm_source=Outcome+Engenuity+2017&utm_campaign=b65aaaa5b5-WWB_Intro_2017_02_08&utm_medium=email&utm_term=0_d9af1bc6e6-b65aaaa5b5-87992581

Matthews, R. (2003). Study of Federal Aviation Administration ramp accident data. Paper presented at *International Society of Air Safety Investigators (ISASI) Seminar*. Washington, DC, USA: International Civil Aviation Organisation.

Matthews, S. (2012). *Safety on the airport ramp: Human factors and management responsibility*. Alexandria, VA, USA: Flight Safety Foundation.

Maurino, D. E. (2000). Human factors and aviation safety: what the industry has, what the industry needs *Ergonomics, 43*(7), 952-959.

Maurino, D. E. (2001). Proactive safety culture: Do we need human factors? In B. J. Hayward & A. R. Lowe (Eds.), *Aviation resource management, Vol. 1* (pp.1-10). Aldershot: Ashgate Publishing.

Maurino, D. (2002). LOSA now Central Focus of ICAO's flight safety and human factors programme. *International Civil Aviation Organisation Journal, 57*(4), 5.

Maurino, D. E. (2008). Threat and error management, line operations safety audit, y normal operations safety survey: La perspectiva de la OACI. Presentation at *Conferencia ASPA/OACI sobre TEM/LOSA/NOSS*. Mexico City, Mexico: International Civil Aviation Organisation & ASPA.

Maurino, D. E., Reason, J. T., Johnston, N., & Lee, R. B. (1995). *Beyond aviation human factors*. Aldershot, UK: Ashgate Publishing.

McDonald, A., & Garrigan, B., Kanse, L. (2007). Confidential observations of rail safety (CORS): An adaptation of line operations safety audit (LOSA). In J. M. Anca (Ed.) *Multimodal safety management and human factors: Crossing the borders of medical, aviation, road, and rail industries* (pp.179-187). Aldershot, UK: Ashgate Publishing.

McDonald, A., Garrigan, B. (2008). Threat and error management: A cross industry comparison. Presented at the *Eighth International Symposium of the AAvPA Conference*. Sydney, NSW, Australia: Australian Aviation Psychology Association.

McDonald, N. & Fuller, R. (1994). The management of safety on the airport ramp. In N. McDonald, R. Fuller & N. Johnston (Eds.), *Aviation psychology in practice* (pp. 68-86). Aldershot, UK: Avebury Technical/Ashgate Publishing.

McDonald, N. (1999). *Aircraft Maintenance: Aircraft Dispatch And Maintenance Safety (ADAMS)* (1996-1999). Amsterdam, The Netherlands, National Aerospace Laboratory.

McDonald, N. (2001). Human systems and aircraft maintenance. *Air Transport, Air & Space Europe, 3*(3/4), 221-224.

McDonald, N. (2001). *SCARF: Safety Courses for Airport Ramp Functions*. Dublin, Ireland: Trinity College. Retrieved on August 23 2014 from: http://www.tcd.ie/psychology/aprg/scarf.html.

McDonald, N. (2002). The development and evaluation of SCARF – and what happened after. *Paper presented at Managing Risk: The Flight Ground Safety Interface*. Melbourne, VIC, Australia: Human Factors Training and Tools.

McDonald, N. (2008). Challenges facing resilience engineering as a theoretical and practical project. In E. Hollnagel, F. Pieri & E. Rigaud (Eds.), *Proceedings of the Third Resilience Engineering Symposium* (pp. 205-210). Juan-les-Pins, France: Presses des Mines.

McDonald, N., Corrigan, S., Daly, C., Cromie, S. (2000). Safety management systems and safety culture in aircraft maintenance organisations. *Safety Science, 34*, 151-176.

McDonald, N., Cromie, S. D., Ward, M. (1997). The impact of safety training on safety climate and attitudes. In H. Soekkha (Ed.), *Aviation safety* (pp. 649-660). Utrecht, The Netherlands: VSP.

McDonald, N., Fuller, R. (1994). The management of safety on the airport ramp. In N. Johnston, N. McDonald & R. Fuller (Eds.) *Aviation Psychology in Practice*. Aldershot, UK: Avebury.

McDonald, N., Fuller, R. (1996). *Safety in airport ground operations*. A proposal to the SCARF International Group to establish a company. Dublin, Ireland: Trinity College.

McDonald, N., White, G., & Cromie, S. (1994). The IATA ramp safety group/APRG ramp accident database. Report to the *Joint IATA Ramp Safety Group/Aerospace Psychology Research group meeting*. Dublin, Ireland: Aerospace Psychology Research Group.

Mearns, K., Flin, R., Gordon, R., Fleming, M. (1997). *Organisational and human factors in offshore safety*. OTH 543 Report. Suffolk, UK: HSE publications.

Mearns, K., Flin, R., Gordon, R., Fleming, M. (1998). Measuring safety climate on offshore installations. *Work and Stress, 12*(3), 238-254.

Mearns, K., Whitaker, S. M., Flin, R. (2003). Safety climate, safety management practice and safety performance in offshore environments. *Safety Science, 41*(8), 641-680.

Merritt, A., Klinect, J. (2006). *Defensive flying for pilots: an introduction to threat and error management*. The University of Texas Human Factors Research Project. Austin, TX, USA: The LOSA Collaborative.

Mitchell, R. J., Williamson, A. M., O'Connor, R. (2009). The development of an evaluation framework for injury surveillance systems. *BMC Public Health, 9*(1), 260.

Montijn, C., Balk, A. D. (2010). *ASC IT – An aviation safety culture inquiry tool – Development from theory to a practical tool*. NLR Technical Report 2009-241. Amsterdam, The Netherlands: NLR.

Montijn, C., De Jong, H. (2006). *Safety culture in air transport: Definition, characteristics, indicators and classification scheme*. Memorandum ATSF-2006-15. Amsterdam, The Netherlands: National Aerospace Laboratory NLR.

Mumaw, R. J., Roth, E. M., Vicente, K. J., Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. *Human Factors, 42*(1), 36-55.

Nascimento, F. A. C., Majumdar, A., Ochieng, W. Y. (2014). Helicopter accident analysis. *Journal of Navigation, 67*(1), 145-161.

National Offshore Petroleum Safety and Environmental Management Authority (NOPSEMA). (2014). *Safety Case Assessment* (Reference No. N-04300-PL0052). Perth, WA, Australia: WA Government.

Nault, R. (2006). The challenge for safety ramp. Proceedings of *the 18th FAA-ATA International Symposium of Human Factors Maintenance and Ramp Safety.* Orlando, FL, USA: Federal Aviation Administration.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3), 263-280.

Negroni, C. (2009). *The most dangerous workplace*. Retrieved on December 22, 2009 from: http://christinenegroni.blogspot.com.au/2009/12/most-dangerous-workplace.html

Newquist Tolk, J. J Cantu J., and Beruvides M. (2015) High Reliability Organization Research: A Literature Review for Health Care, *Engineering Management Journal*, 27:4, 218-237

Nunnally, J. C. (1978). *Psychometric Theory* (pp. 86-113, 190-255). New York, NY, USA: McGraw-Hill.

O'Connor, P., Flin, R. (2003). Crew resource management training for offshore oil production teams. *Safety Science*, *41*(7), 591-609.

O'Connor, P., Flin, R., Fletcher, G. (2002). Methods used to evaluate the effectiveness of CRM training: a literature review. *Journal of Human Factors and Aerospace Safety*, *2*(3), 217–234.

Occupational Safety and Health Administration (OSHA). (2012). Employer safety incentive and disincentive policies and practices. Washington, DC, USA: US Department of Labour: Fairfax, R. E. Retrieved on October 14, 2014, from: https://www.osha.gov/as/opa/whistleblowermemo.html

O'Connor, P., Campbell, J., Newon, J., Melton, J., Salas, E., Wilson, K. A. (2008). Crew resource management training effectiveness: a meta-analysis and some critical needs. *International Journal of Aviation Psychology, 18*(4), 353-368.

O'Hare, D. (2000). The 'wheel of misfortune': a taxonomic approach to human factors in accident investigation and analysis in aviation and other complex systems. *Ergonomics, 43*(12), 2001-2019.

O'Neil, Patrick D. (2012) Policy and Organizational Change in the Federal Aviation Administration: The Ontogenesis of a High-Reliability Organization, Krane University of Nebraska at Omaha Public Administration Review January February 2012

Øien, K., Utne, I. B., Herrera, I. A. (2010). Building safety indicators: part 1: Theoretical foundation. *Safety Science, 49*(2), 148–161.

O'Leary, M., Chappell, S. L. (1996). Confidential incident reporting systems create vital awareness of safety problems. *ICAO Journal, 51*(8), 11–13 & 27.

Ornellas T., (2015) “Twenty Years of Ground handling” *Ground Handling International Magazine* June 2015 page 1 ISSN Number: 1364 – 8330

Orasanu, J., Martin, J., Davison, J. (2001). Sources of decision error in aviation. In G. Klein & E. Salas (Eds.), *Applications of naturalistic decision making* (pp. 3-20). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Oxstrand, J., Sylvander, C. (2010). Resilience engineering: fancy talk for safety culture: A Nordic perspective on resilience engineering. In *3rd International Symposium on Resilient Control Systems (ISRCS)* (pp. 135 – 137), 10-12 Aug, 2010. Idaho Falls, ID, USA: IEEE.

Passenier, D, Sharpanskykh, A. and J. de Boer R."When to STAMP? A case study in aircraft ground handling services." *Procedia Engineering* 128 (2015): 35-43

Patankar, M. S. (2003). A study of safety culture at an aviation organization. *International Journal of Applied Aviation Studies, 3*(2), 243-258.

Patankar, M. S., Taylor, J. (2004). *Applied human factors in aviation maintenance*. Aldershot, UK: Ashgate Publishing.

Patankar, M. S., Taylor, J. C. (2004). *Risk management and error reduction in aviation maintenance*. Aldershot, UK: Ashgate Publishing.

Perneger, T. V. (2005). The Swiss cheese model of safety incidents: are there holes in the metaphor?. *BMC Health Services Research, 5*(1), 71.

Perrow, C. (1984). *Normal accident theory, Living with high risk technology*. New York, NY, USA: Basic Books.

Perrow, C. (1999). *Normal accidents: Living with high-risk technologies*. Princeton, NJ, USA: Princeton University Press.

Petersen, D. C. (1978). *Techniques of safety management* (2nd ed.). New York, NY, USA: McGraw Hill Kogakusha.

Piotrowicz, M., Edkins, G., Pfister, P. (2002). The monitoring of safe behaviour in ramp operations. Paper presented at *the 5th International Human Error and System Safety Development HESSD Workshop*. Newcastle, NSW, Australia.

Pierobon M, "The Challenges Of Detecting Composite Damage, Aviation Pros, Aircraft, Ground Support Worldwide Sept 24, 2014 available from: http://www.aviationpros.com/article/11666312/the-challenges-of-detecting-composite-damage

Rail Safety Regulators' Panel. (2007). *Guidelines for rail resource management*. Publication of the States of Victoria and New South Wales. Melbourne/Sydney, Australia: PTSV/ITSRR.

Rankin, W., Carlyon, B. (2015). Assessing the safety of ramp and maintenance operations. Boeing Commercial Magazine, 2015 Quarter 1, 11. Retrieved from: http://www.boeing.com/commercial/aeromagazine/articles/2012_q2/pdfs/AERO_2012 q2.pdf

Rankin, W., Hibit, R., Allen, J., Sargent, R. (2000). Development and evaluation of the Maintenance Error Decision Aid (MEDA) process. *International Journal of Industrial Ergonomics, 26*(2), 261-276.

Rasmussen, J. (1980). What can be learned from human error reports?. In K. D. Duncan, M. M. Gruneberg, & D. Wallis (Eds.), *Proceedings of an international conference on changes in the nature and quality of working life* (pp. 97-113). Chichester, UK: John Wiley & Sons.

Rasmussen, J. (1982). Human errors: a taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents, 4*(2-4), 311-333.

Rasmussen, J. & Svedung. X. (2000). Proactive Risk Management in a Dynamic Society. Karl, Sweden: Swedish Rescue Services Agency.

Reason, J. (1990). *Human error*. Cambridge, UK: Cambridge University Press.

Reason, J. (1990). The contribution of latent human failures to the breakdown of complex systems. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 327*(1241), 475-484.

Reason, J. (1995). A systems approach to organizational error. *Ergonomics, 38*(8), 1708-1721.

Reason, J. (1997). *Managing the risks of organizational accidents*. Aldershot, UK: Ashgate Publishing.

Reason, J. (1998). Achieving a safe culture: theory and practice. *Work & Stress, 12*(3), 293-306.

Reason, J. T., Hollnagel, E., Paries, J., Drougul, F. (2006*). Revisiting the 'Swiss Cheese' model of accidents.* (EUROCONTROL EEC Note No. 13/06 Project Safbuild). Paris, France: Eurocontol Experimental Centre.

Reason J (2008) The Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries, Ashgate ISBN 9780754674023

Reason, J., Hollnagel, E., Paries, J. (2006). Revisiting the "Swiss cheese" model of accidents. *Journal of Clinical Engineering, 27*, 110-115.

Reason, J., Shotton, R., Wagenaar, W. A., Hudson, P. T. W., Groeneweg, J. (1989). *Tripod: a principled basis for safer operations*. The Hague, The Netherlands: Shell Internationale Petroleum Maatschappij (Exploration and Production).

Reid, J. B. (1970). Reliability assessment of observational data: a possible methodological problem. *Child Development, 41*(4), 1143-1150.

Reimana, T., Rollenhagen, C., Pietikäinen, E., Heikkilä, J. (2015). Principles of adaptive management in complex safety-critical organizations. *Safety Science, 71,* 80–92.

Reynard, W. D. (1991). *The acquisition and use of incident data: Investigating accidents before they happen*. Washington, DC, USA: National Aeronautics and Space Administration: ASRS.

Roberts, K. H. (1990). Some characteristics of high-reliability organizations. *Organization Science, 1*(2), 160-177.

Rochlin, G. I. (1993). Defining high-reliability organizations in practice: A taxonomic prolegomenon. In K. H. Roberts (Ed.), *New challenges to understanding organizations* (pp. 11 – 32). New York, NY, USA: Macmillan.

Rollenhagen, B., Reimana, T. (2011). Human and organizational biases affecting the management of safety. *Reliability Engineering and System Safety, 96*(10), 1263–1274.

Runyan, C. W. (1998). Using the Haddon matrix: introducing the third dimension. *Injury Prevention, 4*(4), 302–307.

Runyan, C. W. (2003). Back to the future: revisiting Haddon's conceptualization of injury epidemiology and prevention. *Epidemiology Review, 25*, 60-64.

Runyan, C. W., Baker, S. P. (2009). Preventing injuries by understanding energy damage. *Bulletin of the World Health Organisation, 87*(5), 402-403.

Salas, E., Burke, C. S., Bowers, C. A., Wilson, K. A. (2001). Team training in the skies: does crew resource management (CRM) training work? *Human Factors, 43*(4), 641-674.

Salas, I., Wilson, K. A., Burke, C. S., Wightman, D. C. (2000). Does crew resource management training work? An update, an extension and some critical needs. *Human Factors, 48*(2), 392-412.

Schwartz, B. M. (1999). Air cargo ground handlers: Is anybody looking? *Transportation & Distribution, 40*, 51–54.

Seagull, F. J., Sanderson, P. M. (2001). Anaesthesia alarms in context: an observational study. *Human Factors, 43*(1), 66-78.

Seidman, I. (2012). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (3rd ed.). New York, NY, USA: Teachers college press.

Sevdalis, N., Davis, R. E., Koutantji, M., Undre, S., Darzi, A., Vincent, C. A. (2008). Reliability of a revised NOTECHS scale for use in surgical teams. *American Journal of Surgery, 196*(2), 184-90.

Sexton, J. B., Thomas, E. J., Helmreich, R. L. (2000). Error, stress, and teamwork in medicine and aviation: cross sectional surveys. *British Medical Journal, 320*(7237), 745-749.

Shappell, S. A., Detwiler, C., Holcomb, K., Hackworth, C., Boquet, A., Wiegmann, D. A. (2007). Human error and commercial aviation accidents: An analysis using the human factors analysis and classification system. *Human Factors, 49*(2), 227-242.

Shappell, S. A., Wiegmann, D. A. (1997). A human error approach to accident investigation: the taxonomy of unsafe operations. *The International Journal of Aviation Psychology, 7*(4), 269-291.

Shappell, S. A., Wiegmann, D. A. (2000a). *The human factors analysis and classification system – HFACS* (No. DOT/FAA/AM-00/7). Oklahoma City, OK, USA: FAA Civil Aeromedical Institute.

Shappell, S. A., Wiegmann, D. A. (2000b). Is proficiency eroding among US naval aircrews? A quantitative analysis using the human factors analysis and classification system. Paper presented at the *44th Annual Meeting of the Human Factors and Ergonomics Society*. San Diego, CA, USA.

Shappell, S. A., Wiegmann, D. A. (2003). Reshaping the way we look at general aviation accidents using the human factors analysis and classification system. Paper presented at the *Twelfth International Symposium on Aviation Psychology*. Dayton, OH, USA: The Ohio State University.

Shay L.A. (2016) "United Airlines Uses Data Visualisation to improve Safety , *Aviation Week &Space Technology* – Inside MRO Aug 11, 2016 –(provided courtesy of Lisa Crocket personal communication 2017)

Shirley, R. B., Smidts, C., Li, M., Gupta, A. (2015). Technique for human error rate prediction (THERP). *Annals of Nuclear Energy, 77*, 194-211.

Shorrock, S., Kirwan, B. (2002). Development and application of a human error identification tool for air traffic control. *Applied Ergonomics, 33*(4), 319–336.

Shorrock, S., Young, M., Faulkner, J. (2005, January/February). Who moved my (Swiss) cheese? *Aircraft & Aerospace,* 31-33.

Shorrock S (2016), Presentation : "*Work-as-Imagined & Work-as-Done: Mind the Gap*" Presented at the Velocity conference- building resilient systems 7/11/2016 Amsterdam, The Netherlands, available from https://conferences.oreilly.com/velocity/devops-web-performance-eu/public/schedule/detail/56268

Silva, S. Lima, M. L., Baptista, C. (2004). OSCI: An organizational and safety climate inventory. *Safety Science, 42*(3), 205 -220

Stanton, N. A, Hedge, A., Brookhuis, K., Salas, E. & Hendrik, H. W. (eds.) (2004). Handbook of human factors and ergonomics methods. Boca Raton, FL, USA: CRC Press.

Stanton, N. A., Ashleigh, M. J. (2000). A field study of team working in a new human supervisory control system. *Ergonomics, 43*, 1190–1209.

Stanton, N. A., Salmon, P., Harris, D., Marshall , A., Demagalski, J., Young, M. S., Waldmann, T., Dekker, S. (2009). Predicting pilot error: Testing a new methodology and a multi-methods and analysts approach. *Applied Ergonomics, 40*(3), 464–471.

Stanton, N. A., Salmon, P., Harris, D., Marshall, A., Demagalski, J., Young, M. S., Dekker, S. (2009). Predicting pilot error: Testing a new methodology and a multi-methods and analysts approach. *Applied ergonomics, 40*(3), 464-471.

Steckel, R. (2014). *Developing and establishing the validity and reliability of the perceptions toward Aviation Safety Action Program (ASAP) and Line Operations Safety Audit (LOSA) questionnaires*. ProQuest Dissertations and Theses; Thesis (Ph.D.) – Saint Louis University; Publication Number: AAT 3631306; ISBN: 9781321097375. Dissertation Abstracts International, Volume: 75-11(E) Section: B., 116.

Stellman, J. M. (1998). Encyclopaedia of occupational health and safety, accidents and safety management (4th ed.). Geneva, Switzerland: International Labour Organization.

Swain, A. D., Guttman, H. E. (1983). *A handbook of human reliability analysis with emphasis on nuclear power plant application*. Washington, DC, USA: Nureg/CR-1278 USNRC.

Swuste, P., van Gulijk, C., Zwaard, A. W. (2010). Safety metaphors and theories, a review of the occupational safety literature of the US, UK and The Netherlands, till the first part of the 20th century. *Safety Science, 48*(8), 1000-1018.

Swuste, P., Van Gulijk, C., Zwaard, W., Oostendorp, Y. (2010). Ocupational safety theories, models and metaphors in the three decades since World War II, in the United States, Britain and the Netherlands: a literature review. *Safety Science, 62*, 16–27.

Thomas M. J. W., Petrilli, R. M., Dawson, D. (2004). *An exploratory study of error detection processes during normal line operations* (Doctoral dissertation). Retrieved from Centre for Applied Behavioural Science.

Thomas, E. J, Sexton, J. B, Helmreich, R. L. (2004). Translating teamwork behaviours from aviation to healthcare: Development of behavioural markers for neonatal resuscitation'. *Quality and Safety in Health Care, 13*(suppl.1), i57-i64.

Thomas, E. J, Sexton, J. B, Lasky, R. E., Helmreich, R. L., Crandell, D. S., Tyson, J. (2006). Teamwork and quality during neonatal care in the delivery room. *Journal Perinatal, 26*(3), 163-169.

Thomas, M. J. W. (2003). Improving organisational safety through the integrated evaluation of operational and training performance: An adaptation of the line operations safety audit (LOSA) methodology. *Human Factors and Aerospace Safety, 3*(1), 25-45.

Thomas, M. J. W. (2003). Operational fidelity in simulation-based training: The use of data from threat and error management analysis in instructional systems design. In *Proceedings of SimTecT2003: Simulation Conference* (pp.91-95). Adelaide, SA, Australia.

Thomas, M. J. W. (2004). Predictors of threat and error management: Identification of core nontechnical skills and implications for training systems design. *The International Journal of Aviation Psychology, 14*(2), 207–231.

Trist, E. (1981). The evolution of socio-technical systems. *Occasional paper,2*, 1981.

Tumulty, B. (2002, April 22). Airline workers at top of list for ergonomic injuries. *The Salt Lake Tribute*. Retrieved from http://www.stlrib.com2002/apr/04222002/business/730389.htm

Turner, B. (1978). *Made-man disaster: The failure of foresight*. London, UK: Butterworth-Heinmann

US Government Accountability Office (US GAO). (2007) Aviation Runway and Ramp Safety :Sustained Efforts to Address Leadership, Technology, and Other Challenges Needed to Reduce Accidents and Incidents GAO-08-29: Published: Nov 20, 2007. Publicly Released: Dec 5, 2007

United States Government Accountability Office (US GAO). (2011). *Aviation safety status of FAA's actions to oversee the safety of composite airplanes* (Doc. No. GAO-11-849). Washington, DC, USA: Us Government Printing Office.

Van de Voordea, P., Monsieurs, K. G., Perkins, G. D., Castrend, M. (2014). Looking over the wall: Using a Haddon Matrix to guide public policy making on the problem of sudden cardiac arrest. *Resuscitation, 85*(5), 602–605.

Vandal, B. (2004). Equipment damage and human injury on the Apron – Is it a cost of doing business? Presented at the *International Society of Air Safety Investigators (ISASI) conference*. Gold Coast, QLD, Australia: ISASI. Retrieved from: http://www.asasi.org/papers/2004/Vandel_Ramp%20Damage_ISASI04.pdf

Vaughan, D. (1996). *The challenger launch decision: Risky technology, culture, and deviance at NASA*. Chicago, IL, USA: The University of Chicago Press.

Veilette, P. R. (2008). Line Operations Safety Audits. *Business & Commercial Aviation, 102*, 32-39.

Verschoor, M., Young, A. (2011). *Willingness to report occurrences in civil aviation in the Netherlands*. The Hague, the Netherlands.

Vincent, C, Amalberti R, (2016) Approaches to Safety: One Size Does Not Fit All, Chapter from *Safer Healthcare* pp 27-37 Date: 14 January 2016

Vincoli, J.W., (1994) Basic guide to accident investigation and loss control. Vol. 1. 1994: John Wiley & Sons

Von Thaden, T. L., Gibbon, A. M. (July, 2008). *The Safety Culture Indicator Scale Measurement System (SCISMS)*. Office of Aviation Research and Development. Washington, DC, USA.

Von Thaden, T. L., Wiegmann, D. A., Mitchell, A. A., Sharma, G., Zhang, H. (2003). Safety culture in a regional airline: Results from a commercial aviation safety survey. In *Proceedings of the 12th international symposium on aviation psychology* (Vol. 139). Dayton, OH, USA: Ohio State University.

Walker, M. (2003). Reasoning with the Reason model. Presented at the Australian Society of Air Safety Investigators (ASASI) Conference. Maroochydore, QLD, Australia: ATSB.

Walker, M.B., and. Bills K.M (2008) "Analysis, causality and proof in safety investigations." *ATSB Transport Safety Report, Aviation Research and Analysis Report* AR-2007-053 (2008).

Ward, M., McDonald, A., Morrison, R., Gaynor, D., Nugent, T. (2010). HILAS project aerospace psychology group. *Ergonomics, 53*(2), 247-267.

Waring, A. E., Glendon, A. I. (1998). *Management, Risk and Change*. London, UK: Chapman & Hall.

Weener, E. (2007). Ground accident prevention – The foundations answer. Paper presented at the *EASS Conference*. Amsterdam, the Netherlands.

Weick, K. E. & Sutcliffe, K. M. (2001). *Managing the unexpected*. San Francisco, CA, USA: Jossey-Bass.

Weick, K. E., Sutcliffe, K. M. (2007). Managing the unexpected: Resilient performance in an age of uncertainty (2nd ed.). San Francisco, CA, USA: Jossey-Bass.

Weick, K. E., Sutcliffe, K. M., Obstfeld, D. (1999). Organizing for high reliability: Processes of collective mindfulness. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 21, pp. 81-123). Greenwich, CT, USA: JAI Press.

Weick, K., Sutcliffe, K. (2001). *Managing the unexpected: Assuring high performance in an age of complexity* (1st ed.). San Francisco, CA, USA: Jossey-Bass.

Weiss, R. S. (1995). *Learning from strangers: The art and method of qualitative interview studies*. New York, NY, USA: Simon and Schuster.

Wenner, C. A., Drury, C. G. (1998). Analysing human error in aircraft ground damage incidents. *Industrial Ergonomics*, 26(2), 177–199.

Wenner, C. L., Drury, C. G. (1996). A unified incident reporting system for maintenance facilities. In *Human Factors in Aviation Maintenance – Phase VI: Progress Report Vol. II* (pp. 191-242). Washington, DC, USA: Federal Aviation Administration/Office of Aviation Medicine.

Werfelman, L. (June 25th, 2009). Ground check. *Aerosafety World*: Flight Safety Foundation (FSF), International Air Transport Association (IATA). Retrievable from: http://flightsafety.org/archives-and-resources/ground-accident-prevention-gap

Werner, C., Drury, C. G. (2000). Analysing human error in aircraft ground damage incidents. *International Journal of Industrial Engineering, 26*(2), 177-199.

Westrum, R. (2006). A typology of resilience situations. In E. Hollnagel, D.D. Woods, & N. Levenson, (Eds.), *Resilience engineering: Concepts and precepts* (pp. 55-66). Burlington, VT, USA: Ashgate Publishing Ltd.

Whitefield, A., Wilson, F., Dowell, J. (1991). A framework for human factors evaluation. *Behaviour & Information Technology, 10*(1), 65-79.

Wickens, C. D., Flach, J. M. (1988). Information processing, in E. L. Wiener & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 111-155). New York, NY, USA: Academic Press.

Wickens, C.D. (2001). Attention to safety and the psychology of surprise. Keynote address at the *Eleventh Symposium on Aviation Psychology*. Columbus, OH, USA: Ohio State University.

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

Wiegmann, D. A., & Shappell, S. A. (2001). Applying the Human Factors Analysis and Classification System (HFACS) to the analysis of commercial aviation accident data. Paper presented at the *Eleventh International Symposium on Aviation Psychology*. Columbus, OH, USA: Ohio State University.

Wiegmann, D. A., & Shappell, S. A. (2003). *A human error approach to aviation accident analysis: The Human Factors Analysis and Classification System*. Bodmin, Cornwall, UK: Ashgate Publishing Ltd.

Wiegmann, D. A., Zhang, D., Von Thaden, T. L., Sharma, G., and Mitchell A. A. (2002). *A synthesis of safety culture and safety climate research* (Technical Report ARL-02-3/FAA-02-2). Washington, DC, USA: Federal Aviation Administration (Contract DTFA 01-G-015).

Wiegmann, D., von Thaden, T. L. (2007). A review of safety culture theory and its potential application to traffic safety. In D. Nelson, and B. Tefft (Eds.), *Safety Culture and Traffic Safety*. Washington, DC, USA: AAA Foundation Publishing.

Williams, J. C. (1988). A data-based method for assessing and reducing human error to improve operational performance. Paper presented at the *Fourth IEEE Conference on Human Factors in Nuclear Power Plants*. Monterey, CA, USA: IEEE.

Williamson, A. M., Feyer, A. M., Cairns, D., Biancotti, D. (1997). The development of a measure of safety climate: the role of safety perceptions and attitudes. *Safety Science, 25*, 1527.

Wilson, J. R. (Ed.). (2005). *Rail human factors: Supporting the integrated railway*. Aldershot, UK: Ashgate Publishing Ltd.

Woods, D. B., Cook, R. I. (2004). Mistaking error. In B. J. Youngberg & M. J. Hatlie (Eds.) *The patient safety handbook* (pp. 95–108). Sudbury, MA, USA: Jones & Bartlett.

Woods, D. D., L. J. Johannesen, R.I. Cook, and N.B. Sarter, eds. (1994). *Behind human error: Cognitive systems, computers and hindsight* (No. CSERIAC-SOAR-94-01). Dayton, OH, USA: Dayton University Research Institute (URDI).

Woods, D., Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. Boca Raton, FL, USA: CRC Press / Taylor & Francis.

WorkSafe Victoria. (2012). *Hazard identification, risk assessment & risk control in the workplace*. Retrieved on 4 Jun, 2015 at: www.worksafe.vic.gov.au/hazard-identification-risk-assessment-&-risk-control-in-the-workplace.html.

Wu, C. L. (2008). Monitoring aircraft turnaround operations – framework development, application and implications for airline operations. *Transportation Planning and Technology, 31*(2), 215-228.

Zhang, H., Wiegmann, D. A., von Thaden, T. L., Sharma, G., Mitchell, A. A. (2002). Safety culture: a concept in chaos?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 46, No. 15, pp. 1404-1408). SAGE Publications.

Zohar, D. (1980). Safety climate in industrial organisations: theoretical and applied implications. *Journal of Applied Psychology, 65*(1), 96- 102.