January 2019

Can You Hear Me, Major Tom? Open Issues In Extra-Vehicular Activity Communications

Elizabeth Howell

CAN YOU HEAR ME, MAJOR TOM? OPEN ISSUES IN EXTRA-VEHICULAR ACTIVITY COMMUNICATIONS

by

Elizabeth Howell
Bachelor of Journalism, Carleton University, 2007
Master of Science, University of North Dakota, 2012

A Dissertation
Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of
Doctorate of Philosophy

Grand Forks, North Dakota

August
2019

This dissertation, submitted by Elizabeth Howell in partial fulfillment of the requirements for the Degree of Doctorate in Aerospace Studies from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

James Casler

Richard Fiordo

Kimberly Kenville

Pablo De Leon

Thomas Petros

This dissertation is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Christopher Nelson

Associate Dean of the School of Graduate Studies

7/8/19

Date

PERMISSION

Title	Can You Hear Me, Major Tom? Open Issues In Extra-Vehicular Activity Communications
Department	Space Studies
Degree	Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Elizabeth Howell

August 2019

TABLE OF CONTENTS

TABLE OF FIGURES

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to my adviser, Dr. James Casler, and the rest of my advisory committee for their support and advice during my time at the Doctorate of Philosophy program at the University of North Dakota. Dr. Casler and his wife in particular were quite generous with their time, opening their homes and their lives to me several times between 2013 and 2019 when I visited Grand Forks for my residency requirements. Committee member Pablo De Leon guided me through the application process when applying to the University of North Dakota's IRB. Committee member Kimberly Kenville suggested a useful track for my meta-synthesis. I also appreciated the support of committee members Richard Fiordo, Thomas Petros, and (now-retired) Warren Jensen, whose feedback on the dissertation helped improve it.

Thank you also to the hard-working analog crews at the ILMAH who recorded several hours of video for transcription, which was no easy task as they needed to modify the recording setup and the simulated spacesuits to accommodate the recording.

Administrator Bev Fetter at the Department of Space Studies is a tireless student advocate, even at the busiest times of the term. I could not have done this without her help. Distance librarian Janet Rex provided support in searching for papers for the meta-synthesis. In general, the resources of the Space Studies Department and Chester Fritz Library were incredibly useful for tracking down even obscure space papers, especially through the interlibrary loan service. I am very grateful for this help.

Some of my friends and colleagues willingly pitched in to help with the harder parts of the dissertation. Transcriptionist Patti Moran enthusiastically listed to many hours of noisy, technical talk during the EVAs and made constructive suggestions about how to improve future recordings. Virtual assistant Christina Goodvin generated the references for my meta-analysis into APA format, on short notice. Coders Kayla La France, Chris Nicholson and Kam Yee took time out of their busy schedules to do approximately 40 hours of work for each transcript set, reading and analyzing them to back up my methodology. Countless friends cheered me on through the process, in particular Beth (who earned a Ph.D. not too long ago), Ashley and Evan. You folks are the best.

Thank you also to many family members. My sister and brother both have or had connections at Carleton University in Ottawa, and they willingly borrowed books from the library as requested. I also must acknowledge the excellent library at the University of Ottawa, where I taught periodically during my Ph.D. My parents generously offered to pick up part of my travel costs whenever I visited Grand Forks. My in-laws – including my sister-in-law, who is doing a dissertation herself – provided needed moral support.

A final thanks goes to my husband, J, for all of his love during this process. J and I met shortly after I began my Ph.D., and I don't think either of us really knew what a dissertation really entailed until I started getting into the thick of the research. But he learned quickly and was incredibly supportive, even when I doubted myself. Thanks to you, I'm a better person and researcher.

For J

Who will take me to infinity – and beyond

ABSTRACT

High-reliability organizations (HRO) and organizations in isolated, confined environments (ICE) both operate under conditions where reliability is expected, but do not appear to have similar emphases placed on total reliability, based on a brief survey of the literature. A content analysis searched out a stronger relationship between HRO and ICE. Leadership and team size are hypothesized as differences between HRO and ICE, since the literature appears to show HRO as taking place in larger teams with more distinct hierarchies. This dissertation examined this postulation, based on two sub-hypotheses. Hypothesis 1 is that the error rate of a team's actions is inversely related to the size of the team, based on the distinction between HRO and ICE. Hypothesis 2 is that transformational leaders in ICE reduce the number of human errors compared with transactional leaders, since Bass suggests transformational leaders better inspire their teams to improve. Two datasets were gathered to test these hypotheses. The first, in support of Hypothesis 1, was a meta-synthesis of team literature. The second, in support of Hypothesis 2, were new recordings of extra-vehicular activities (EVAs) from two crews at the University of North Dakota Inflatable Lunar/Martian Analog Habitat (ILMAH). The result for Hypothesis 1 is inconclusive, and the result for Hypothesis 2 was rejected.

CHAPTER 1

INTRODUCTION

Reliability background

This paper seeks to better define reliability, with reference to how it is performed in isolated environments. Mission reliability is a key concept in space exploration. One definition of reliability states that it is "the quality of being trustworthy, or of performing consistently well." (Oxford University Press, 2018).[1] Astronauts participating in lengthy off-Earth missions are concerned with reliability of systems, especially in locations where it is difficult to communicate with Mission Control. A potential approach to increasing reliability for these missions lies in the intersection of teams operating in isolated, confined environments (ICE) – which include the space environment – and high-reliability organizations (HRO). There may not be as high an emphasis on reliability in ICE as there is within an HRO. Understanding why this is, and whether this is an issue for astronauts, could be a way of improving reliability procedures for operations in ICE environments more generally.

While it is difficult to define ICE precisely – different researchers have different expectations of how isolated and extreme the environment is – what is agreed upon is that the environment is dangerous, going home is difficult or impossible, and that crews must be mainly reliant upon themselves to achieve their work. (Golden, Chang & Kozlowski, 2018; Sandal, Leon & Palinkas, 2006.) ICE has variously been used to describe underwater or Antarctic bases (Palinkas & Anderson, 2003; Kanas et al., 2010), prisons (Suedfeld & Steel, 2000), spaceflight (Suedfeld, Brcic & Legkaia, 2009; Jacubowski et al., 2015) or exploring remote areas (Leon,

1 This definition of mission reliability should not be confused with statistical reliability, which is defined as: "the degree to which the result of a measurement, calculation, or specification can be depended on to be accurate." (*Ibid*).

Sandal & Larsen, 2011). Some researchers describe these missions also as being monotonous (Bishop, Grobler & Schjoll, 2001).

Clearly, reliability is a priority in ICE. Simply put, ICE environments are defined as regions with little to no external support, meaning a single mistake there can lead to catastrophic consequences. Indeed, performance issues are reported as common in long-duration isolated conditions, such as in Antarctica, leading authors to point out a need for better studies concerning how to improve human performance. There are many facets under study, including psychological issues, interpersonal conflict, expectations for privacy and on the social side, how the crew is organized and led. (Stuster, 1996, 4). While reliability is addressed to some degree in ICE literature, there is less discussion than is found in HRO papers about reliability aspects such as accountability, wariness, error-reporting and frequency of errors.

HRO was first defined during studies of reliable organizations in the 1980s, at the University of California, Berkeley, principally by researchers Karlene Roberts, Gene Rochlin and Todd LaPorte. A paper from that era states the research interest arose because a "small group of organizations" in American society can do highly technical tasks daily without failing. As examples of HRO, the authors cite utility grid management, air traffic control, and flight operations aboard U.S. Navy aircraft carriers. (Rochlin, La Porte & Roberts, 1987). Some of the characteristics of high reliability organizations include paying attention to emergent problems (Rampersand, & Rampersand, 2007; Brass, Olney, Glimp, Lemaire, & Kingston, 2018), using failures as an opportunity to implement "lessons learned" (Christianson, Sutcliffe, Miller & Iwashyna, 2011) , and using procedures and routines to deal with a rapidly changing environment (Roberts, 1990; McIver, Lengnick-Hall, Lengnick-Hall, & Ramachandran, 2013.)

In terms of reliability, multiple publications emphasize the importance for leadership to ensure safety, although the mechanisms by which leaders can enforce reliability are still poorly understood. (Martínez-Córcoles, 2018, 237). While it is simplistic to blame failures solely on human error – error is one larger part of systems theory that examine why incidents occur – understanding why errors occur does improve the reliability of systems when the "root causes" of those mistakes are uncovered, whether those errors are due to leadership, badly designed systems or other causes (Wickens, Lee, Liu & Becker, 2004, 367-369; Dekker, 2006, 15). However, much of the literature on error-making and leadership comes from conditions that are not isolated, e.g., factories or large organizations. This leaves open questions about how leaders lead in ICE, and reduce errors. Paradoxically, in ICE the role of leadership may be diminished because of the isolated condition in itself. At least one meta-analysis showed that the unusual experience of living together in close quarters and under harsh conditions may make it more difficult for leaders to assert their rank and to make decisions based on status, because hierarchies in these locations tend to be more informal. However, prevalence of mistakes was not quantified. (Stuster, *ibid*, 96). There are, however, many different influences that could affect the rate of error-making in ICE teams, including the impact of isolation, confinement or the extreme environments in which teams operate.

Team sizes of three to six people are common both in spaceflight and in analog ICE environments, in part due to the size constraints placed upon crews by available crew size in habitats and the limits of environmental supports such as air or water (Ark, Sipes, Holland & Cockrell, 2010; Duchesne & Tressler, 2010; Suedfeld & Steel, 2000). If crew sizes diminish below a certain threshold, crew goals cannot be fully achieved due to maintenance requirements of the facility. This is what happened with a 2018 HI-SEAS team that ceased operations in part due to physical and psychology issues leading to several team members leaving their posts (Koren, 2018.)

A 1997 meta-analysis showed that in organizations, group size as well as composition determine overall team performance and satisfaction (Furnham, 1997.) In ICE, however, success of a team may not only depend on size, but also on individual crew member's personalities, specifically the ability to cope in challenging circumstances. (Sandal, Bergan, Warncke, Vaernes & Ursin, 1996; Sandal, 2000). Certain functions in ICE may also be offloaded to an offsite team. Mission Control in Houston, for example, controls the mission planning, consumption of fuel and oxygen, and other facets so that NASA crews may focus on orbital tasks (Kraft, 2001, 100). It is especially important for a larger organization to create a culture of safety, because otherwise even the best-performing team may overlook issues in favor of the organization's direction (Padgett, Gossett, Mayer, Chien & Turner, 2017; Gaw, Rosinia & Diller, 2018).

It is unclear whether leadership influences teams in ICE, among the other conditions cited. Moreover, if leaders do impact ICE teams, a related question is whether leaders can reduce errors in those teams despite the unique challenges of the environment. This dissertation explores two hypotheses concerning leaders, and the teams that they govern. To explore the possible impact of ICE on team errors, Hypothesis 1 states that the error rate of a team's actions is inversely related to the size of the team; as the team size increases, error diminishes. This is because the literature suggests that HROs are part of larger teams than ICE, so accordingly, it may be that HROs have more team members and can thus concentrate more fully on reliability. For example, a team working in a nuclear power plant or on an aircraft carrier may number in the dozens or more, while ICE teams tend to be limited to much smaller sizes of under 10 people, and commonly in spaceflight, between three to six people.

As will be explored later in this paper, Bass (1990) is a key reference in describing leadership styles, and in his highly cited work, he points out that transformational leaders inspire their teams to go beyond the minimum requirements – something that a transactional leader fails to do. Paradoxically, although a transactional leader focuses on avoiding mistakes, a transformational leader may do better from a productivity perspective because their actions focus on the good of the group – thus inspiring people to work harder and do actions to help the team. (*Ibid*). While Bass does not state this explicitly, perhaps this relates to reliability as well, although an alternative explanation for leadership in ICE could better be explained by distributed leadership based on task expertise, rather than ranking. Thus, Hypothesis 2 states that transformational leaders in ICE reduce the number of human errors compared with transactional leaders in ICE.

Reducing errors background

Better understanding the nature of reliability will assist future space explorers with planning their missions, especially if – as NASA desires – crews in the coming decades venture to the Moon and to Mars. These environments are much more distant from planet Earth than the International Space Station (ISS). In the case of Mars, crews will work in an environment where their radio communications will take several minutes to travel to Earth. Crews will thus need to work more autonomously, without necessarily relying on Mission Control to the same extent that ISS crews do today. This may include (but not necessarily) reducing errors, since human errors are often included as a factor in investigations of aerospace incidents. The literature disagrees to what extent humans should be implicated in these "normal accidents", meaning incidents arising out of the interface of multiple complex systems (Perrow, 1984) – more details about normal accidents will be explored later in this dissertation. Nevertheless, implementing procedures such as checklists demonstrably reduces errors. By extension, checklists also reduce the possibility of incidents not only in aerospace, but other challenging environments such as medicine, construction and high-stakes business negotiations. (Gawande, 2009)

This dissertation assumes that reliability is related to error-making, and examines the phenomenon of error-making through two independent studies: a meta-synthesis of published errors in team performance, and field studies of two crews working in a simulated ICE environment, namely the UND Inflatable Lunar/Martian Analog Habitat (ILMAH.) The rationale for each of these investigations will be explored in the forthcoming chapters. The goal is to provide guidance on reducing errors in a space analog environment. These interim results could be expanded or adapted to other studies to help other crews increase their reliability in challenging circumstances.

It is acknowledged (and shown later in this dissertation) that error-making is a complex matter that includes many factors. Leadership and organizational size were chosen as possible factors based on the fact that HROs are part of larger organizations compared with ICE, although more study may be required to better establish that link. This dissertation is a small step towards that goal.

Purpose of the study and research questions

The purpose of this study is to find a way to reduce the number or frequency of errors committed in ICE environments, through examining two possibilities for implementing a reduction – team size and leadership style. Specifically, the questions to be answered were: Does team size play a role in the number or frequency of errors? Does leadership style play a role in the number or frequency of errors?

In ICE, team size is necessarily small because in isolated conditions, the size of habitats and the complications of travel make it difficult for large teams to participate in ICE. With the exception of Antarctica's McMurdo Station, teams in ICE are limited in size. In most environments outside of isolated conditions, more people can always arrive on site to bolster the size of the effort, even in situations such as forest rescues. Hypothesis 1 is that the error rate of a team's actions in ICE is inversely related to the size of the team.

As a first step to see whether leadership has any influence at all on error-making, the principal investigator examined two different styles of leadership to see what differences in error-making may be seen. As will be explained in the Literature Review, there are various approaches to leadership, including the Full Range Leadership model incorporating transactional, transformational, laissez-faire and many other examples. Transactional and transformational leadership styles were selected, however, due to the depth of literature available contrasting their characteristics, and also because well-cited sources such as Bass (1990) considered these leadership styles' influences on error-making. Thus, Hypothesis 2 is that transformational leaders in ICE reduce the number of human errors compared with transactional leaders in ICE.

CHAPTER II

REVIEW OF THE LITERATURE

Isolated, Confined Environments (ICE)

In the literature, space exploration – or more precisely speaking, working in a small environment such as a space station or a spacecraft – is considered an example of ICE. There are certainly degrees of ICE, even within different space programs. For example, a crew venturing out to the moon during the Apollo program of the 1960s and 1970s faced more distant isolation challenges over a short period, while a current International Space Station crew is close to Earth but facing isolation challenges over a long period. While the degree of ICE can be argued among different space programs, what unites these various environments is distance from the usual support systems of work and family, the difficulty of transportation to and from the space environment, the lack of a familiar physical setting, and a crew that must be largely self-reliant to accomplish tasks. Space crews are therefore extensively trained in the tasks that are required to carry out a successful mission. Astronauts also receive extensive psychological support from NASA that may also be qualified as training. As with other ICE environments, astronauts also perform tasks on a tight schedule. Tasks in space can range from running science experiments, to doing repairs, to live press conferences, to performing individual procedures during complex EVAs. (National Aeronautics and Space Administration, 2004).

Individual astronauts face psychological stressors in ICE. In the Antarctic, these problems can include balancing work and leisure, accommodating individual preferences for food and hygiene, and group dynamics in a confined space. (Stutster, *ibid*, 21) A literature review of analog and spaceflight environments (Geuna, Brunelli & Perino, *ibid*), as well as observations of crew members in orbit (Kanas, 1998; Kanas, 2014) yielded external stressors such as privacy and leadership, and internal stressors such as health problems, loss of energy and changes in behavior. These psychological dynamics may also play into leadership.

The text *Bold Endeavors* is considered one of the definitive sources on ICE, although its focus is on polar expeditions and not so much on spaceflight. Author Jack Stuster cites Capt. Paul D. Nelson, a Navy social psychologist who studied the characteristics of successful leaders at U.S. Antarctic stations. He found that groups accord status and esteem to those leaders who (a) consult with specialists and individuals for technical/task-based decisions; (b) consult with the entire group before making a decision that affects the group as a whole; (c) make emergency decisions quickly and if needed, autocratically. The challenge of leadership is balancing among these situations, Stuster added, with intelligence, interpersonal skills and motivation. Other studies, he added, show that leadership in an isolated environment becomes more respected with time, as the leader accumulates experience. (Stuster, 2011, 100)

Another meta-analysis by Peter Suedfeld, who has studied many space crews and space analog crews, suggests that leadership in ICE is fraught with difficulty due to factors such as few status markers, few gatekeepers such as secretaries to manage a leader's time, a lack of support from the organization sponsoring the organization, and an abundance of experts on the crew who do not feel uncomfortable in challenging the leader. (Suedfeld, 2010) A survey of space leaders (Suedfeld, 2008) suggested that mission commanders tend to concern themselves more with the morale of the group than individual tasks. He mused in a later study (Suedfeld, 2010) that such a finding about space leadership was "interesting", but "compatible" with expedition literature in Antarctica because of the importance of interpersonal relations and emotions in ICE environments.

However, a study by noted Antarctic research Lawrence Palinkas – citing other work – shows that task leadership is more preferable during the early stages of a mission, when the camp is being set up, and supportive leadership is preferable late in the mission (Gunderson & Nelson, 1963; Nelson, 1964, Palinkas, 2001), presumably when the crew needs more psychological assistance and time. Preferences for leadership style may also depend on the personality of the individual crew member, although the literature shows varying positive and negative results for that relationship in a submarine environment and in an online survey of a Nordic company (Sandal, Endresen, Vaernes & Ursin, 1999; Hetland, Sandal & Johnsen, 2008.)

Another study notes that expedition tensions or conflicts can in part be attributed to poor or ineffective leadership, competition between leaders, or even difficulties between leaders and followers (Palinkas & Suedfeld, 2008); a later study by Palinkas suggests that the ideal leader "should possess both task-oriented and supportive-oriented leadership traits" (Palinkas, 2001), particularly based on a 135-day Mir simulator study that showed high crew cohesion in association with the leaders' emphasize on task leadership – as well as personality, which included the ability to support (Kanas, Weiss & Marmar, 1996).

These findings may be less applicable to international crews, however, as findings from aviation and analog environments show that there are differing values in various countries concerning hierarchical leadership, as well as following rules and procedures – favorite techniques of transactional leaders. (Helmreich & Merritt, 1998; Helmreich, 2000). Indeed, communication problems such as misunderstandings, language barriers and work style are "often mentioned challenges of multi-national space crews", not only because of national culture differences, but also because of the different ways in which different national space organizations are structured. (Sandal & Manzey, 2009)

The link between leadership and error-making in ICE is less clear, but there is some literature discussing why errors occur. With regard to error in ICE, Stuster (*ibid*) cites case studies from space and Antarctica using examples such as sleep disturbances (46), fatigue (76-

77), "exceeding an individual's capacity for workload" (81) and low workload (81). One remedy he suggests is having not only good communication from leaders, but also (borrowing from aviation) communicating in a standard way that reduces errors. (168-9). While Stuster does not specifically link the common practice of aviation/space exploration checklists to reducing errors, he does mention checklists in other parts of the book, advocating for their use where it would make sense. Aviation uses checklists as part of standard operation procedures (SOPs) to ensure a high degree of safety in routine and non-routine operations.

Unfortunately, there is little peer-reviewed recent literature that looks specifically at the role of errors in ICE, presumably because there are so few participants. A search of " 'isolated confined environment' and error" yielded just nine papers on the University of North Dakota Chester Fritz library database during a January 2018 search. One paper's citation led to another paper discussing the role of error in space analogs, which noted a strong cultural correlation concerning relationships between leaders and followers, and in adhering to rules and procedures. The paper (which took data from 26 nations) also elucidated five types of errors in analogs: intentional non-compliance errors, procedural errors, communication errors, proficiency errors and operational decision errors. (Helmreich, 2000). The papers at Chester Fritz also noted there is a perception that error is frequent in ICE (Palinkas, 2001); that error is linked to poor communication (Suedfeld, 2003); that error is linked to sleep disturbances, high workload or psychosomatic discomfort (Kanas et al., 2009); or that there is decrease in accurate physical performance (e.g., pointing) due to microgravity (Geuna, Brunelli & Perino, 1995). Four papers were discarded as they only mentioned errors in relation to the authors' statistical analysis.

Also, a search of the NASA Technical Report Server yielded four ICE papers in relation to error. One source found that error is linked to monotony on long space missions, or to exhaustion (Slack et al., 2016). Another study presented a preliminary effort to search for signs of cognitive load using EEG measurements. The author stated that personnel with a "transient cognitive impairment" (due to illness, medication, intoxication, fatigue or other factors) "may be error prone in situations that tax the limits of their reduced mental capacity." (Gevins, Smith, McEvoy & Brown, 1999). A third study described multiple NASA efforts to combat error-making using human factors principles (e.g., combating fatigue and overwork) that are similar to previous studies earlier in this dissertation. (National Space Biomedical Research Institute Annual Report, 2000). This search shows that NASA researching the impact of errors in ICE, although it appears the measures to do so are measures that have been proposed in other literature on error-making.

Leadership in ICE

ICE is a difficult context in which leaders must operate, and a handful of studies do suggest there could be an effect of context on leadership (Lord, Foti & De Vader, 1984), although more study is needed to know this for sure. Authors determined several factors that could separate out leaders by context: national culture (Brodbeck et al., 2000; Koopman et al., 1999), hierarchical leader level, and environmental characteristics such as dynamic versus stable (Brown & Lord, 2001; Keller, 1999; Lord et al., 2001), and how much conformity individuals have in a group, e.g., the military (Mischel, 1977). The latter two characteristics – environmental and conformity – would have special application to ICE, because many of these environments are militaristic in nature and are also extremely dangerous or isolated. In such a situation, surrounded by few resources and in a structure that encourages conformity, leaders may ultimately have a choice: to adapt to their circumstances in order to carry out their mission, or to stick to a single agenda at the risk of being the "odd one out" in the group.

In ICE, authors such as Stuster lean heavily towards a more contingency-based leadership that is adaptable to the circumstances. Similarly, another team states that transformational leaders (which are discussed later in this Literature Review) are more likely to challenge or look for new methods that oppose existing operations, and prefer risk-based opportunities. In their words, "[t]ransformational leaders do not merely react to environmental circumstances -- they attempt to shape and create them." (Lowe, Kroeck & Sivasubramaniam, 1996)

However, there are some examples of contextual leadership that can be cited. One study of a range of ICE team performance – supplemented with data collection from several isolated missions with application to spaceflight (e.g., NASA Extreme Environment MIssion Operations or NEEMO, Human Exploration Research Analog or HERA, and Antarctica) – points out that team composition "should be considered as a foundational context for understanding how a team is likely to work together, even when membership is unchangeable." (Salas et al., 2015) A study of a four-subject, 180-day mission in China showed that leadership styles varied depending on the context: "The task and supportive roles are not independent, both roles are important for group cohesion and mission success and are used in conjunction over the course of isolation and confinement." (Wu, Quianying, Xiong, Xu & Li, 2018). Antarctic groups have also shown a preference for task leadership near the beginning, and supportive leadership near the end (Stuster, 1996; Leon, Sandal & Larsen, 2011), which may necessitate a shift in leadership style by any managers.

High Reliability Organizations (HRO)

As stated earlier in this paper, an HRO is an organization that does difficult and highly technical tasks daily with few or no errors, such as air traffic control or utility grid management.

The organizational and social behavior of HRO, as well as its reliability, stand in contrast to "normal accident" theory, outlined in detail by Charles Perrow around the same time period that HRO studies began. Perrow argued that “normal accidents” are “an integral characteristic of the system” in situations of complexity and tight coupling. (Perrow, 1984, 5) In spaceflight, an example of a situation of complexity and tight coupling could be the launch of a rocket. Later research on "normal accidents" theory, however, argues that disasters actually "arise from an absence of some kind of knowledge at some point" and may not necessarily be a function of the system. (Turner & Pidgeon, 1997, 3) The disagreement between HROs and normal accident theory persists for much of the literature in the 1980s and 1990s. At times, proponents of HRO argued that normal accident theory was incomplete or did not exist, and vice-versa.

Intersections with ICE may be elucidated by studying HROs that took place in hostile or isolated environments. Examinations of isolated environments studied for HRO include submarines (Bierly & Spender, 1995; Bierly, Gallagher & Spender, 2008) mountaineering (Allard-Poesi & Giordano, 2015), polar expeditions (Burke, Shuffler & Wiese, 2018), long-duration sailboat racing (*ibid*) and spaceflight, although authors generally say NASA cannot be considered an HRO (Boin & Schulman, 2008; Burke, Shuffler & Wiese, 2018; Casler, 2014; LaPorte, 2006) with the exception of McCurdy, 1993, who argues NASA is a "culture supported exceptionally high levels of performance for tasks very difficult to perform." Examples of hostile environments studied include naval aviation (Ciavarelli & Crowson, 2004), precision armies – which were said not to be HROs (Demchak, 1996), nuclear power plants (LaPorte, 1982; Dietrich & Childress, 2004; Offstein, Kniphuisen, Bichy & Childers, 2013), aircraft carriers (Laporte & Consolini, 1998; Roberts, 1990), oil refineries (Lekka & Sugden, 2011) and nuclear weapons maintenance (Tolk & Hartley, 2010).

It appears that the leadership styles for HRO in these isolated and hostile environments are by no means fixed, perhaps because the leadership style must be adaptive to the circumstances – much as Stuster (*ibid*) described in the Antarctic. (This may also be said of many business organizations operating outside of ICE, although examining a typical corporation falls outside the scope of the dissertation). At least one paper advocates using two leaders (where possible) to work together in a distributed leadership model, because different leadership functions such as task assignment and motivation may naturally fall to different leaders. (Allard-Poesi & Giordano, 2015). Another alternative may be using a collection of experts for each component or platform of a system, which may be required due to the system's complexity or isolation (Bierly, Gallagher & Spender, 2008; Burke & Shuffler & Weise, 2018); there is debate as to whether the leadership structure is flexible as the situation requires, so that certain experts may come to the fore in a dynamic scenario (Demchak, 1996) or whether leadership is best based as a hierarchical style where one person is the ultimate voice in making decisions (LaPorte & Consolini, 1998; Offstein, Kniphuisen, Bichy & Childers, 2013).

Leaders and error-making

The relationship between leaders and error-making is unclear in HRO when it comes to hostile or isolated environments. These papers state (as the other HRO literature states) that HROs focus on avoiding error (e.g., Lekka & Sugden, 2011). Thus if one assumes error-making is something with which everyone concerns themselves, perhaps the leader's goal is more transformational than transactional (e.g., Boin & Schulman, 2008). In ICE, at least one author calls for having an external leader to help with safety, because they "may be in the best position to facilitate transition functions due to the ability to maintain a broader situation awareness that is sometimes constrained within teams under pressure." (Burke, Shuffler & Wiese, 2018).

The mission design of ILMAH includes brief missions of about one to two weeks in length, in a single type of isolated condition that simulates spaceflight. Thus, it is clear that a single study on the ILMAH cannot state definitively whether a particular leadership style or particular rate of errors is indicative of an HRO in an isolated or hostile condition, because the literature varies considerably. The possibility can exist, however, because HROs have been used to describe teams in spaceflight. What remains to be seen is whether HROs in spaceflight can be translated to analog environments, such as ILMAH. This would be an interesting direction for future studies.

With application to spaceflight, it is unclear whether NASA and other space agencies as a whole are HRO. NASA, for the most part, traditionally operates under a bureaucratic structure (Hacker & Grimwood, 1977), which may be opposite to the aims of an HRO. The agency has a strong military background that arose in the 1960s through the U.S. Air Force, as demonstrated by the German engineers who created the Mercury, Apollo and Gemini rockets. NASA also followed the Air Force practice of contracting to private firms for most of its hardware needs. (*Ibid*.)

Although an HRO may not encompass an entire public organization, there may be aspects within that organization that are an HRO. For example, while the U.S. Navy as a whole would not be considered an HRO, it has been argued that aircraft carriers operated by the Navy are. Submarines may also be considered an HRO environment. On submarines, crews de-emphasize experiential learning due to the potential for catastrophe. Instead, incident reports and crew rotation updates provide opportunities for crew members to examine critical events and to seek opportunities for improvement. (Bierly, Gallagher & Spender, 2008.)

Errors

Before measuring errors in ILMAH, it is useful to know how the literature classifies errors more generally. One source suggests errors can be divided into three main types: slips and lapses (memory failures, misattention, habit intrusions, and the like); rules-based mistakes (e.g., misapplying a rule in a given situation) and knowledge-based mistakes (which occurs in novel situations) (Reason, 2008). There are at least a few limitations of this list. The first is that it does not take into account the design of a system, such as its displays or warnings. In relation to leadership, however, the more relevant issue is the list does not necessarily account for mistakes made in groups, such as groupthink. The Challenger shuttle explosion of 1986, for example, is blamed in part on production pressure and "normalization of deviance" when it came to certain anomalies, including faulty O-rings in boosters. (Vaughan, 1996, xiii-xiv) These decisions were the responsibilities of groups of engineers and senior management, who conferred before the launch, and cannot be traced back to any single person. Still, the culture of "bad apples" is so rampant in engineering that more than 30 years after the fact, individual engineers such as Bob Ebeling still blamed themselves (and not the group as a whole) for the disaster. (Berkes, 2016)

A related view to normal accident theory is the hypothesis that human error is an offshoot of complexity. This is in contrast to blaming "bad apples" for causing accidents. As one author argues, "The occasional human contribution to failure occurs because complex systems need an overwhelming human contribution for their safety." (Dekker, 2006, 65.)

Many experts have instead argued that complex systems are best described under systems theory, which was first proposed by biologist Ludwig von Bertalanffy in the 1940s and refined by Rosh Ashby in the 1950s. Simply put, systems theory focuses on the relations between individual subsystems, and also takes into account the environment in which the subsystems operate, including the influence of outside operators. (Heylighen & Joslyn, 1992). A related hypothesis – called STAMP (Systems-Theoretic Accident Model and Processes) – focuses on how to stabilize systems by utilizing feedback loops of information and control. The system is understood as a dynamic process. If an accident occurs, it is because the feedback did not adjust sufficiently to performance changes to maintain safety. (Leveson, 2004).

The contribution of groups is better elucidated in an introductory text to human factors engineering. Citing several papers from the 1980s and the 1990s, the chapter discusses possible influences such as training, group size, individual personalities and stress. Key positive influences of leaders include setting a vision, embodying specific performance goals, getting individual commitments from team members to work together, and advocating for co-ordination and shared accountability. (Wickens, Lee, Liu & Becker, 2004.)

Systems theory and leadership are easier to study in large organizations, because these organizations are common in the business world. A leader in ICE, however, must deal with a complex system with a small group. While support systems are commonly available, they are remotely located and may not be easy to contact quickly in the case of an emergency.

Leadership in small groups

Small group studies for ICE are hindered by many problems. As stated already, sample size remains a large issue for ICE due to how few people have participated over the years, particularly in spaceflight. The literature also tends to be biased towards groups of males, providing little insight about how isolation affects other genders.

Even the definition of a small group reveals limitations. One literature review suggests it is probably about 20 members or less, but adds that most "small group" studies concern five people or less. (Shaw, 1981, 6). For context, a typical ISS crew ranges between three and six people. Most small-group investigations appear to be done in the laboratory, making it difficult to generalize what is seen there to a less controlled environment. (Seal, Bogart & Ehrnhardt, 1998.) It also appears to be difficult to correlate results across studies. A literature review of 100 examinations concerning non-verbal communication in small groups, for example, concluded that there are too many questions posed to draw a meaningful consensus. (Gatica-Perez, 2009.)

Leadership in ICE may be informed by questions about how leadership is shaped in small groups. It could depend on how often a person speaks in a group (Maricchiolo, Livi, Bonaituo & Gnisci, 2011); whether leaders function better as efficient autocrats or collaborative diplomats (Prahl, Dexter, Braun & Van Swol, 2013); how leadership changes in digital vs. “real-world” environments (Tromp, Bullock, Steed, Sadagic, Slater & Frecon, 1998) and whether leaders function best if they act similarly to the group (Shaw, 1981, 15-16).

Scientific investigations of leadership – in Europe and North America – began around the turn of the twentieth century, focusing on "great man and trait" approaches; through the 1940s and beyond, leadership understanding evolved into contingency approaches (Hunt, 1999), which moves closer to the leadership styles that are considered for this dissertation. The roots of contingency leadership emerged from two seminal studies in the 1970s: (House, 1977) and transformational leadership (Burns, 1978). Both studies heavily influenced the first seminal text on transformational and charismatic leaders, called *Leadership Beyond Expectations* (Bass, 1985). Models of leadership subsequently split into several different schools, whose history has been well-covered in other literature reviews. Here, we will focus on the origins and arguments concerning the full-range leadership model that also emerged during the late 1970s.

Perhaps the first study of full range leadership (Burns, 1978) emerged in relation to studies of two United States presidents, Franklin Delano Roosevelt and John F. Kennedy. Burns argued that these presidents were successful because they exhibited both transformational and transactional leadership. They were transformational in that they could lead the public with inspiring words, yet transactional because they gained votes through political deals. Burns argued that these two characteristics – transactional and transformational leadership – lie on a continuum. Leadership styles are commonly evaluated today using the Multifactor Leadership Questionnaire (MLQ), although it has been criticized for emphasizing transformational leadership (Yammarino & Dubinsky, 1994; Kelloway, Barling & Helleur, 2000) and for being weak in measuring transactional and non-leadership (Hinkin & Schriesheim, 2008).

Here it is useful to discuss transactional leadership and transformational leadership in more detail. A highly cited paper by Bass (1990), which is among the first to discuss this concept in the academic literature, defines these leadership styles as follows. Transformational leaders "broaden and elevate the interests of their employees ... generate awareness and acceptance of the purposes and mission of the group, and ... stir their employees to look beyond their own self-interest for the good of the group" (p. 21). Transactional leaders "explain what is required of them [employees] and what compensation they will receive if they fulfill these requirements" (pp. 19-20). The characteristics of these leadership types are outlined in Table 1 below; note that more recent research often places laissez-faire into its own category separate from transactional leadership (Eagly, Johannesen-Schmidt & van Engen, 2003; Chaudhry & Javed, 2012).

Table 1: Characteristics of transformational and transactional leaders (Bass, 1990)

Transformational	**Transactional**
Charisma: Provides vision and sense of mission, instills pride, gains respect and trust.	Contingent reward: Contracts exchange of rewards for effort, promises rewards for good performance, recognizes accomplishments.
Inspiration: Communicates high expectations, uses symbols to focus efforts, expresses important purposes in simple ways.	Management by Exception (active): Watches and searches for deviations from rules and standards, takes corrective action.
Intellectual Stimulation: Promotes intelligence, rationality and careful problem solving.	Management by Exception (passive): Intervenes only if standards are not met.
Individualized consideration: Gives personal attention, treats each employee individually, coaches, advises.	Laissez Faire: Abdicates responsibilities, avoids making decisions.

The debate between transactional and transformational leadership is a common one in business literature, but some authors also argue for a third type of leadership – transformative – that appears to be emerging in educational and social justice literature. Shields (2010) writes that there have been some studies of transformative leadership in recent years, but they use different definitions (e.g. Quantz, Rogers & Dantley, 1991; Weiner, 2003; Hoffman & Burrello, 2004). Following her own interviews with educators, she defines transformative leaders as such: "Transformative leadership begins with questions of justice and democracy; it critiques inequitable practices and offers the promise not only of greater individual achievement, but of a better life lived in common with others." Her paper, which was in the context of educational practices, continues: "Transformative leadership, therefore, inextricably links education and educational leadership with the wider social context within which it is embedded." (Shields, *ibid*, 559). Shields defines the differences between transactional, transformational and transformative leadership in the table below. The scale is shown below in Table 2. Bass, for his part, has expanded his leadership theory into a full-range leadership model that includes elements of transformational, transactional and laissez-faire management. (Avolio and Bass, 1991).

Table 2: Transactional, transformational, transformative leadership (Shields, 2010)

	Transactional	Transformational	Transformative
Starting point	A desired agreement or item	Need for the organization to run smoothly and efficiently	Material realities and disparities outside the organization that impinge the success of individuals, groups and organization as a whole
Foundation	An exchange	Meet the needs of complex and diverse systems	Critique and promise
Emphasis	Means	Organization	Deep and equitable change in social conditions
Processes	Immediate cooperation through mutual agreement and benefit	Understanding of organizational culture; setting directions, developing people, redesigning the organization, and managing the instructional program	Deconstruction and reconstruction of social/cultural knowledge frameworks that generate inequity, acknowledgment of power and privilege, dialectic between individual and social
Key values	Honesty, responsibility, fairness and honoring commitments	Liberty, justice, equality	Liberation, emancipation, democracy, equity, justice
Goal	Agreement; mutual goal advancement	Organizational change; effectiveness	Individual, organizational, societal transformation
Power	Mostly ignored	Inspirational	Positional, hegemonic, tool for oppression as well as action
Leader	Ensures smooth and efficient organizational operation through transactions	Looks for motive, develops common purpose, focuses on organizational goals	Lives with tension and challenge; requires moral courage, activism
Related theories	Bureaucratic leadership, scientific management	School effectiveness, school reform, school improvement, instructional leadership	Critical theories (race, gender), cultural and social reproduction, leadership for social justice

CHAPTER III

METHOD

Setting

While there are more than 60 years of general team research available (Golden, Chang & Kozlowski, 2018), within ICE studies the environment limits the scope. That is because the current standard of research on teams requires using surveys that are cross-sectional, time-lagged or laboratory-based. Within ICE, teams work in conditions far from studying scientists and are rarely examined directly (Marks et al., 2001, Cronin, Weingart & Todorova, 2011; Kozlowski & Chao, 2012). Even space agencies regularly use spaceflight analogs in ICE to study astronaut teams and isolated teams directly, especially because it is imperative to determine the efficacy of individual astronauts working alone and in groups before undertaking an expensive spaceflight. These analogs (which tend to be located in isolated areas, and where crew members practice spaceflight procedures such as spacewalks) focus on testing five main hazards of spaceflight that can be simulated on Earth, including lowered gravity fields, increased isolation and confinement, hostile and closed environments, space radiation and distance from Earth. (NASA, 2018).

Only a portion of the crew's overall mission conversations were captured in the ILMAH recordings, but recordings were limited to the EVA to provide a similar standard of privacy for crew members that astronauts have on the ISS. Specifically, ISS astronauts usually do not have their activities broadcast publicly unless they are participating in a press conference, an educational event with children, a crew handover between expeditions, or a spacewalk. In these ILMAH conversations, it was assumed that the crew members were speaking truthfully and freely and that they were working towards their mission objectives without encumbrances. However, their speech and intentions may have been altered by the Hawthorne effect, or the fact that they were aware of being recorded and observed for the purpose of this investigation. (McCambridge, Witton and Elbourne, 2014)

Data collection and meta-synthesis (Hypothesis 1)

This meta-synthesis was in response to Hypothesis 1, answering the question of whether increasing team size diminishes error rate. The researcher consulted with the UND Chester Fritz Library about appropriate terms to use, since librarians are used to producing appropriate terms for computer databases to understand. With suggestions in hand, the principal investigator performed a search in April 2019, using both the UND Chester Fritz Library (which includes a cross-section of space journals with full text access, available for student researchers) and the NASA Technical Reports Server or NTRS (which is highly cited by researchers within the space industry, and also contains full text references).

An initial search of || "team size" AND (failure* OR error* OR mistake*) || in Chester Fritz's database, as recommended by a UND distance librarian, produced 2,683 results, resulting in the need for more exclusion criteria. It was decided to reframe the search as || "team size" AND (failure* OR error* OR mistake*) AND crew ||. "Crew" is a commonly used term in aviation, military and spaceflight referring to a self-contained group of people who sometimes work remotely (or in extreme environments) from the central hub of their organization. This exclusion criterion reduced the number of possible studies in Chester Fritz to 99.

Within NTRS, the search of || "team size" AND (failure* OR error* OR mistake*) AND crew || produced six results. While the search for || "team size" AND (failure* OR error* OR mistake*) || only produced 12 papers in comparison to Chester Fritz's thousands of results, the NTRS database search revealed several papers on software development which had little or no relevance to crews working in extreme environments.

Thus, the search || "team size" AND (failure* OR error* OR mistake*) AND crew || in Chester Fritz and NTRS produced a total count of 106 papers that were used in the meta-synthesis. This next produced the research question of to what degree ICE and HRO is considered within these results. The following searches using ICE and HRO in Chester Fritz and NTRS produced few or no results:

- Chester Fritz: "team size" AND (failure* OR error* OR mistake*) AND "isolated, confined environment" = 0 results
- Chester Fritz: "team size" AND (failure* OR error* OR mistake*) AND "high reliability organization" = 1 result
- NTRS "team size" AND (failure* OR error* OR mistake*) AND "isolated, confined environment" = 0 results
- NTRS: "team size" AND (failure* OR error* OR mistake*) AND "high reliability organization" = 0 results

It was thus decided not to include pre-selected HRO or ICE papers artificially in the meta-synthesis, for two reasons: (1) on their own, it appears there has been little research performed within HRO and ICE on team sizes and errors (2) there was a risk of precluding the validity of the results of the meta-synthesis, which proceeds by strict inclusion and exclusion criteria. Future researchers may want to hand-search papers in these fields to seek references of team size and mistakes to see whether there are results that are not being captured in these large databases by computer searching. So it should be understood that this set of papers captures a cross-section of literature that may have applicability to HRO and ICE, but does not necessarily include these particular fields.

For a qualitative analysis of the data, the PI proceeded using a meta-synthesis. A quick scan of papers in the Chester Fritz Library suggests that the bulk of meta-syntheses have been used in nursing and other health-care fields, so its efficacy in ICE and HRO so far, remains unknown. This dissertation will thus serve as a test case. Some medical units may be considered HRO, however, so there is justification for using a meta-synthetic approach.

The attributes that distinguish qualitative meta-synthesis (QMS) from meta-analysis are that the meta-analysis addresses quantitative studies (Leary & Walker, 2018) and is a "statistical procedure that attempts to integrate a body of quantitative research." (Erwin, Brotherson & Summers, 2011). The purpose of QMS is to obtain knowledge that is otherwise inaccessible. (Lachal, Revah-Levy, Orri & Moro, 2017). This methodology intends to translate, explain, and discover meaning. (Erwin, Brotherson & Summers, *ibid*.) The emphasis is on interpretation, which will enable the development and refinement of theory (Finlayson & Dixon, 2008) while retaining the uniqueness of individual studies. This interpretation of findings from selected qualitative studies may obtain "richer, more complete understanding of the phenomenon" of interest (Erwin, Brotherson & Summers, *ibid*). An issue is to synthesize qualitative studies to achieve some generality, while preserving the relevance of the individual studies. (*Ibid*). Inclusion decisions based on quality assessments become a source of bias. (Finlayson & Dixon, *ibid*).

Given the number of approaches available, it was decided to use the approach recommended by Finlayson and Dixon, which has been used or cited more than 100 times in other studies. Again, the efficacy in HRO and ICE remains to be demonstrated, but the approach has been peer-reviewed and used in several examples of meta-synthetic studies, principally in nursing and health care (which is true of the rest of meta-syntheses in general). How to proceed with a qualitative analysis is thus not well defined, but Finlayson & Dixon suggest using "formal appraisal criteria or checklists", giving examples such as Burns (1989) and the Critical Appraisal

Skills Program (CASP 1999). These criteria or checklists give the qualitative process more rigor, similar to a quantitative analysis, although as stated earlier there are issues in terms of selection bias and the analysis itself. It was decided to proceed with the CASP checklist since the materials included separate 10-point questionnaires for qualitative studies and literature reviews. These two questionnaires can be easily duplicated with other researchers and thus, the efficacy of the qualitative analysis may be tested in future studies. The CASP website (2019) says it has been used for more than 25 years, principally in the health care field, and a quick search of Google Scholar reveals that CASP has been used in the peer-reviewed literature thousands of times.

Discussion of the findings is contained in Results.

Data collection and field studies (Hypothesis 2)

Field studies at the ILMAH were performed to look at the effects of leadership on error in analog ICE environments for spaceflight, using an analog that is as high-fidelity as possible so that the leadership and error-making could be better argued to have applicability to spaceflight.

While there are numerous space analogs available for researchers in the United States and Canada, ILMAH was chosen for a few considerations. First, UND is practically unique among American universities in that it has its own spacesuit designs being tested in an analog, providing a somewhat high-fidelity simulation of what astronauts experience when working in these bulky outfits. Spacesuits can affect communications because of the need to use radio, and the fact that crew member mobility and visibility is restricted (affecting gestures, facial expressions, and other tools commonly used in communication.) Also, the ILMAH suits are pressurized. This not only adds realism to the activity, but it also reduces natural mobility while increasing stress and fatigue. Pressurized suits are an uncommon tool in analog environments, such as the Mars Desert

Research Station (MDRS) and Hawai'i Space Exploration Analog and Simulation (HI-SEAS).

Second, and on a related note to the spacesuits, ILMAH crew members perform EVAs in a similar fashion to astronauts. Mission objectives for each EVA are determined ahead of time, and the crew members go out for a limited number of hours – inside of the spacesuits – to perform geological or technological tasks in situ. The closest space equivalent to these experiences is the Apollo missions of the 1960s and 1970s, in which astronauts walked upon the moon. The research being performed at ILMAH may also help with determining procedures for Mars missions, should NASA or other agencies decide to send astronauts there.

Third, ILMAH has a separate "mission control", which is a common feature of spaceflights. As with space station crews, members of ILMAH missions must communicate using radio and must wait for a response from mission control while making decisions on the spot. This is comparable to a spaceflight environment, although there may be uncaptured differences (such as scheduling, or the effect of distance between a team and mission control) between the professional astronaut teams in orbit and the amateur analog participants on the ground.

Fourth, ILMAH missions are extremely time-bound – similar to space shuttle missions and missions performed during the Mercury, Gemini and Apollo eras. Findings from ILMAH work may not have as much validity on the International Space Station, where most crews work for at least five months or six months, because issues of crew health and isolation may mount after multiple months in space. However, ILMAH findings can be compared with missions of short duration (a few weeks), when psychological and physical health are not as badly affected. As with crew members in space, ILMAH crew members engage in a variety of science and technology studies and are limited in their spare time.

ILMAH runs missions approximately two to three times a year with three people per crew, allowing for enough mission frequency to generate results in a short time. Pending IRB certification and approval from the ILMAH team, UND graduate students are encouraged to participate in ILMAH experimentation. Other analog facilities accept proposals on a competitive basis, or run missions less frequently.

A brief overview of the ILMAH is useful here. The UND Human Space Flight Laboratory received a NASA grant in 2009 for the North Dakota Planetary Exploration Initiative. The primary goal was to develop, construct and use a habitat to house up to four crew members for as long as 30 days, as well as a Pressurized Electric Rover with two NDX-2 planetary suits externally attached. The University of North Dakota is the first university with a NASA-funded laboratory to research space exploration and planetary surface exploration suits. Crews have been working on the ILMAH since 2013, running missions of between 10 to 30 days each. (UND Aerospace, 2015).

In 2013, NASA awarded UND a $750,000 Experimental Program to Stimulate Competitive Research (EPSCoR) grant. The grant allowed the ILMAH to expand with four new modules: a geology module, an EVA module, an exercise and human performance module, and a greenhouse module. These modules were gradually integrated into missions, starting in 2015. (UND Aerospace Human Spaceflight Laboratory, 2017).

EVA participants were selected through competitive application for each mission, through a process overseen by principal investigator (PI) Dr. Pablo de Leon. The principal investigator of this study did not participate in subject selection. While relationships among the participants were not a direct factor for selection, many of the participants were graduate students at UND. Thus they already were colleagues at the university and knew each other professionally, which may have influenced the results.

As with all UND human research, the ILMAH aligns its work with the university's

Institutional Review Board (IRB) to protect the welfare of participating human subjects. This meant that for the principal investigator's research study, most of the considerations for human research were already covered under an umbrella authorization from the IRB (IRB-201310-121). Crew members had the option to opt out of the data collection at any time, even after their mission concluded.

The principal investigator submitted her own proposal to the IRB requesting passive audio recordings of the participants as they were conducting extra-vehicular activities (EVAs). The initial one year approval was used to collect EVA data, then a one-year extension was used to analyze the data collected and to write the findings in this dissertation.

In space exploration, EVAs are commonly broadcast real-time on venues such as NASA Television, which is available on select cable stations and also worldwide on the Internet. As such, astronauts participating in EVAs traditionally have a lessened expectation of privacy. Crew activities inside ISS are usually not broadcast in real time. Presumably, this is due to privacy concerns, or because the astronauts perform experiments subject to commercial or scientific limitations. NASA, however, does not offer recordings of its EVAs due to restrictions in personnel in NASA Television and related offices.

The principal investigator's EVA study did not define the tasks to be conducted. Rather, it was a passive activity performed during research activities for other IRB-approved investigations. After the EVAs were performed, the principal investigator received copies of crew reports for each EVA outlining the objectives, accomplishments and lessons learned for their work, to assist in the analysis of the transcripts. Some of the tasks the crew performed included deploying a weather balloon, testing simulated emergency procedures, and navigating and driving a rover to designated locations.

A hired transcriptionist and the principal investigator reviewed all of the transcriptions and removed any identifying information on the crew members and participating scientists, such as names and age, to protect subject privacy. Subject designators were applied in place of names of the participating crew members, based on designators provided by the participating crew. ILMAH PI Pablo de Leon gave his consent for his name to be included in the transcriptions, so his information was preserved. Occasionally, the EVA participants mentioned identifying information (such as names) for people who were not participants. These characteristics were removed to protect those people's privacy.

Overall, data were collected for two missions at the ILMAH, which will be defined here as Mission 1 and Mission 2. The numbers used to designate missions in this study do not match up with the internal ILMAH numbering system to designate individual missions, in order to protect subject privacy. The ultimate goal of analyzing the transcripts was to better understand what predominant style of leadership the crew members adopted over the course of the mission. To be clear, "leadership" in this context was understood as task leadership, rather than identifying the assigned leader of a group (i.e., the mission commander). Further discussion of how leadership and error are connected is contained within the Results section.

For the ILMAH analysis, it was decided to use ordinal numbers to best rank the crew members and avoid skewing. To the best of the principal investigator's knowledge, this has not been used in analog studies before, and thus, this is an untested methodology that will need to be examined further in other studies. However, Stevens (1946) suggested that ordinal numbers – the determination of greater or less – is applicable to isotonic groups or regression, or fitting a line to a series of non-linear observations (Stevens, *ibid*.). The PI determined the ILMAH set was indeed non-linear, as shown by the large swings in percentage counts in between Schwartz values. As mentioned before, the data were clustered into zones of extremely low counts and extremely high counts. Ordinal scales allow measurements of differences between individuals, as well as the direction of the difference. However, the size of the difference cannot be measured. (Gravetter & Wallnau, 24).

Here is how ranking in this dissertation was performed: For a given value, such as Achievement, the raw counts were examined. The crew member with the highest number of Achievement values was given a rank of 1, the crew member with the second highest number of Achievement values was given a rank of 2, and the crew member with the third highest number of Achievement values was given a rank of 3.

Once the rankings were established, these were placed against the Schwartz values for every individual leadership style measured in the system: transformational, transformative and transactional. Crew members were then assigned a leadership style based on the number of "1s" they had in each style. The style with the most "1s" was suggested as the dominant leadership style for a given crew member. So for example, if a crew member had received three "1s" in transformational leadership, two "1s" in transactional leadership and one "1" in transformative leadership, they would be designated a transformational leader. Occasionally, a crew member had an equal number of "1s" in different leadership categories, in which case a tie-breaker was employed. The leadership style with the most number of "2s" would be the assigned leadership style. The full results are shown in the Results, in Tables 11 and 12.

Identifying leadership using value scales of Schwartz (Hypothesis 2)

To attempt consistency with past studies, this paper applies a methodology used in previous ICE research, based on a list of value scales recorded by Schwartz (1992). He identified values that are universal to humans in at least 20 countries, according to the initial research study. It is a scale that already has been tested in the literature against first-person accounts of subjects, so it was believed that analysis of transcripts would be a similar enough process for validity. Second, the work of Schwartz is highly cited and regarded in the literature, which adds to the scale's validity overall when considering the validity of measuring human emotions.

The list was used by noted ICE researcher Peter Suedfeld and his associates in multiple research studies examining coping strategies in spaceflight (Suedfeld & Weiszback, 2004; Suedfeld, Brcic & Legkaia, 2009; Suedfeld, Wilk & Cassel, 2014). This scale has also been used in other ICE environments, such as transcripts of cave explorers (MacNeil & Brcic, 2017) or narratives of Holocaust survivors (Suedfeld, Krell, Wiebe & Steel, 1997.) Thematic content analysis has other cited advantages. It can be applied to various data sources (oral, written, archived, "real time", different languages, etc.) and "is totally nonreactive and unobtrusive" (Suedfeld & Weiszback, 2004).

Table 3: Value scale adapted from Schwartz (Suedfeld & Weiszback, 2004)

Value	Meaning
Achievement	Personal success through demonstrated competence according to social standards
Benevolence	Concern for close others in everyday interaction
Conformity	Inhibition of socially disruptive acts, impulses or inclinations
Hedonism	Pleasure in satisfying orgasmic needs
Power	Social prestige/status, control over people and resources
Security	Safety, harmony, stability of society, relationships and self
Self-direction	Independent thought and action: choosing, creating, exploring
Spirituality*	Meaning and harmony by transcending everyday reality (Schwartz said this may not be universal to all cultures)
Stimulation	Excitement, novelty, challenge
Tradition	Respect for one's cultural/religious customs and ideas
Universalism	Understanding, appreciation, tolerance and protection of all the welfare of all people and nature

Although Suedfeld used this value scale to outline astronaut coping strategies, one paper suggests a methodology to map these values on to different leadership styles. Sarid (2016) wrote that the original intention of his study was to "address central conceptual and methodological complications in leadership styles", particularly in the Multifactor Leadership Questionnaire that is often used to measure variants in Multifactor Leadership Theory (which is explored in Literature Review.) The tool also allows researchers to examine the relationship between leadership values/beliefs as well as leadership styles, particularly when talking about questions of ethics. (*Ibid*, 1*.)*. Notably, Sarid did not include hedonism in his tool because it is "viewed by Schwartz to intersect openness and self-enhancement … [making it] unhelpful for the tool presented here." (*Ibid*, 15.)

Sarid linked the values of transformational, transformative and transactional leadership using definitions by Shields (2010) and Brown (2004) and then using the corresponding values in the Schwartz value scale that intersected these definitions. For example, for transformative leadership, Shields defined the following characteristics (Shields, p. 562; Sarid, p. 14):

1. A combination of both critique and promise;
2. Attempts to affect both deep and equitable changes;
3. Deconstruction and reconstruction of the knowledge frameworks that generate inequity;
4. Acknowledgment of power and privilege;
5. Emphasis on both individual achievement and the public good;
6. A focus on liberation, democracy, equity and justice; and
7. Evidence of moral courage and activism.

Subsequent to his seminal 1992 paper, Schwartz defined how each of his values are related to four separate characteristics (2012). These were "openness to change" (self-direction, stimulation, hedonism), "self-enhancement" (hedonism, achievement, power), "conservation" (security, conformity, tradition) and "self-transcendence" (universalism, benevolence). Note that Schwartz did not include "spirituality" among the values, so his subsequent research has reduced the number of values to 10 – not the original 11.

Sarid (2016) defined leadership styles using these four characteristics defined by Schwartz. He argued that transactional leadership was a balance between self-enhancement and conservation, transformational leadership was a balance between self-enhancement and openness to change, and transformative leadership was a balance between openness to change and self-enhancement. He did not include hedonism in his leadership styles as it intersects two of Schwartz's characteristics, making it difficult to judge where it lies on the leadership scale. The relations between Sarid and Schwartz are shown in Table 4.

Table 4: Schwartz scale (1992) compared to three leadership constructs (Sarid, 2016)

<table>
<tr><th>Motivation</th><th>Power</th><th>Achievement</th><th>Universalism</th><th>Benevolence</th><th>Hedonism</th></tr>
<tr><td>Stimulation</td><td colspan="2" rowspan="2">Transformational leadership
(self-enhancement and conservation)</td><td colspan="2" rowspan="2">Transformative leadership
(self-transcendence and openness to change)</td><td rowspan="5">No associated leadership style</td></tr>
<tr><td>Self-direction</td></tr>
<tr><td>Tradition</td><td colspan="2" rowspan="3">Transactional/ instructional leadership
(self-enhancement and openness to change)</td><td colspan="2" rowspan="3"></td></tr>
<tr><td>Conformity</td></tr>
<tr><td>Security</td></tr>
</table>

Finding the definitions of Schwartz's values to be ambiguous, the principal investigator, after using one transcript as a test case, reformulated the definitions against the crew members' interactions. Consistent with Schwartz's leadership construct, spirituality was not used. The following scale was thus applied to the ILMAH transcripts in Table 5:

Table 5: Schwartz's leadership constructs applied to the ILMAH transcripts

Achievement	Celebrating success of a task, e.g., a crew member saying "Good job"
Benevolence	A crew member acknowledges another crew member, e.g., through thanking them, through asking a question, through answering a question
Conformity	A crew member performs an act to appear congruent to the rest of the crew, e.g., joining in with a joke
Hedonism	A crew member is sarcastic or satisfies an orgasmic need (e.g., burping, or making a joke.)
Power	A crew member giving an order, e.g., telling another crew member to drive the rover in a particular direction
Security	Speaking about matters of security or safety, e.g., a crew member losing air from their spacesuit
Self-direction	A crew member making a decision or judgment, e.g., telling Mission Control that they will proceed 50 meters to the next work site, or telling Mission Control that a task is complete
Stimulation	A crew member noting something novel during the mission, e.g., seeing planes in the sky
Tradition	A crew member describing how something is always done, e.g., saying past crews have implemented a procedure and it is best to follow it
Universalism	A crew member referring to a shared experience among the crew members or among humanity, e.g., a movie that everyone has seen

Of note: not all portions of the transcripts were included in the analysis. Identifying information about the subjects, personnel and other people mentioned were removed. Other parts of the recordings were inaudible or garbled, or were obscured due to excessive background noise. Seven transcripts were analyzed for Mission 1, and five for Mission 2. Two transcripts were discarded due to recording issues.

Each transcript analysis was performed by the principal investigator and two supplementary coders. The coders were initially selected among recently graduated UND Space

Studies graduate students with a communications background. It was decided to use former Space Studies students, rather than current ones, to reduce the chance of a coder recognizing one of the participating ILMAH crew members (who were largely drawn from the current UND student population). During the second round of analysis, one of the coders was unavailable due to work obligations. The principal investigator thus selected a replacement coder that she knows professionally, who has a keen amateur interest in spaceflight (including attending launches), as well as a theoretical background that includes a Ph.D. in cognitive science and employment at a technology company that deals in part with communications.

For each mission, the principal investigator and two supplementary coders participated. Each person had a copy of the transcript with identifying information removed. The coders used the comment function in Microsoft Word to do their analysis. Each sentence was coded separately, unless the sentences could be grouped together into a similar theme. Each code included the subject's number, and a designation of Schwartz's value. Sample comments included (without quotes) "117: Self-direction", "458: Benevolence", "385: Security".

Subsequently, the principal investigator counted how many times each Schwartz value appeared for each subject. This was initially performed by doing a simple search of the comments in Microsoft Word and manually counting how many times the value appeared. For later transcripts, the Microsoft Word document was converted into PDF format and transferred to Adobe Reader, featured a capability to automatically count comments.

To determine inter-rater reliability, as a first guide the principal investigator looked to a previous study (Suedfeld, Brcic & Legkaia, 2009) that also used the Schwartz value scale. The PI used Microsoft Excel to calculate a Pearson coefficient to evaluate inter-rater reliability, finding a reliability of 0.86 or higher. (*Ibid.*) (1 is considered perfect reliability). A Pearson correlation – a common measure of reliability among coders – was chosen to map the relationship between two measurement sets. However, correlation does not necessarily mean causation. That is to say, that the variables of interest change relative to each other, but it does not show that one change is the result of another. Also, outlier data points can skew the value correlation. (Gravetter & Wallnau, 519-20). Note that in this particular study, the outlier data points were not modified.

CHAPTER IV

RESULTS

Purpose of the study and description of the sample

ICE literature appears to have a lack of discussion on reliability when compared with HRO literature, based on a brief examination of the literature. The principal investigator has two hypotheses as to why this would be the case. Hypothesis 1 is that the error rate of a team's actions is inversely related to the size of the team, since HROs operate within a larger team framework than ICE. Hypothesis 2 is that transformational leaders in ICE reduce the number of human errors compared with transactional leaders. Below are findings related to each hypothesis, as well as related discussion from the literature.

The sample for Hypothesis 1 included 31 studies obtained as part of a meta-synthesis on team literature. The sample for Hypothesis 2 was two, three-person crews working at the IMLAH. One of those crew members participated in both missions, leaving a final sample size of five people.

Hypothesis 1: Number of errors in small teams

Hypothesis 1 states the error rate of a team's actions is inversely related to the size of the team, and the methodology that was decided to find this was to search out references of team size and error in the literature, as explained in Method. The 106 studies found at Chester Fritz and NTRS, which are listed under Meta-Synthesis in the References section, were then put through a qualitative meta-synthesis approach as recommended by the Critical Assessment Skills Programme (1999), based on a recommendation from Finlayson and Dixon (2008).

The papers were hand-examined by the PI with a few exclusion criteria in mind: (1) any quantitative study, since that is not covered under the CASP methodology (2) any study that was off-topic, in that team size and error (either together, or separately) were not addressed whatsoever. This winnowing process reduced the number of examined papers to 31 papers, which are listed in Table 6 along with the results.

Each qualifying author was placed in an Excel spreadsheet, with the numbers of the question on the horizontal axis and the author's name on the vertical axis. Each CASP question on the worksheet (Table 7) was answered with "yes" or "no", with the number 1 assigned to yes/positive answers and the number 2 assigned to no/negative answers. Table 7 also includes the screening methodology used for this paper to answer "1" or "2" to each question. The PI also added an 11^{th} question – "Does this paper grade or address team size with relation to error?", which is shown in Table 6 along with the other 10 questions from CASP. Papers were eliminated from the analysis as soon as they received a "2", in line with the elimination by aspects methodology, a commonly used behavior for decision-making. The methodology involves identifying most important attribute is identified for each heuristic; any alternatives are eliminated when they fall below that cutoff point. (Goodwin & Wright, 2004, 19).

Table 6: CASP methodology applied to considered papers

Lead author (year)	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Bell (2016)	1	1	1	1	1	1	1	1	2	x	x
Bowers (2000)	1	1	1	1	1	1	1	1	1	1	2
Chakravorty (2010)	1	2	x	x	x	x	x	x	x	x	x
Choi (2010)	1	1	1	1	1	2	x	x	x	x	x
Cohen (1997)	1	1	1	1	2	x	x	x	x	x	x
Courtenay (2013)	1	1	1	1	2	x	x	x	x	x	x
Crichton (2004)	1	1	1	1	1	2	x	x	x	x	x
Christian (2017)	1	1	2	x	x	x	x	x	x	x	x
Day (2004)	1	1	2	x	x	x	x	x	x	x	x
Guzzo (1996)	2	x	x	x	x	x	x	x	x	x	x
Hung (2013)	1	2	x	x	x	x	x	x	x	x	x
Jones (2000)	1	1	2	x	x	x	x	x	x	x	x
Klimoski (1994)	1	2	x	x	x	x	x	x	x	x	x
Langer (2016)	1	1	1	1	1	1	1	1	1	1	2
Lawton (2012)	1	1	1	1	1	2	x	x	x	x	x
Marlow (2018)	1	1	1	1	1	1	1	1	1	1	2
McEwan (2017)	1	1	1	1	1	1	1	1	1	1	1
Mogilever (2018)	1	1	1	1	1	2	x	x	x	x	x
Muller (2015)	1	1	1	1	1	2	x	x	x	x	x
Naikar (2006)	1	1	2	x	x	x	x	x	x	x	x
Ornato (2014)	1	2	x	x	x	x	x	x	x	x	x
Paris (2010)	2	x	x	x	x	x	x	x	x	x	x
Pratoom (2018)	1	1	1	1	1	1	1	1	1	1	2
Propp (2010)	1	1	1	1	1	2	x	x	x	x	x
Ricci (2012)	1	2	x	x	x	x	x	x	x	x	x
Salas (2008)	1	1	1	1	1	1	1	1	1	1	1
Ulrich (2017)	2	x	x	x	x	x	x	x	x	x	x
Usman (2018)	1	1	1	1	1	2	x	x	x	x	x
Weaver (2010)	1	1	1	1	2	x	x	x	x	x	x
Weiss (2016)	1	2	x	x	x	x	x	x	x	x	x
Wilson (2007)	1	1	2	x	x	x	x	x	x	x	x

Table 7: CASP questionnaire

Literature Review	Qualitative Study
1. Did the review address a clearly focused question? *Answered 1 if the question was easily answerable through a yes or no question*	1. Was there a clear statement of the aims of the research? *Answered 1 if the question was easily answerable through a yes or no question*
2. Did the authors look for the right type of papers? *Answered 1 if a systematic literature search was performed, including listings of keywords and databases*	2. Is a qualitative methodology appropriate? *Answered 1 if the methodology best used surveys or interviews, as opposed to quantitative measures*
3. Do you think all the important, relevant studies were included? *Answered 1 if a systematic literature search was performed, including listings of keywords and databases*	3. Was the research design appropriate to address the aims of the research? *Answered 1 if the methodology best used surveys or interviews, as opposed to quantitative measures*
4. Did the review's authors do enough to assess quality of the included studies? *Answered 1 if the authors discussed the merits and drawbacks of each paper*	4. Was the recruitment strategy appropriate to the aims of the research? *Answered 1 if the local population used in the study addressed the research question*
5. If the results of the review have been combined, was it reasonable to do so? *Answered 1 if the authors combined papers and explained their similarities and differences*	5. Was the data collected in a way that addressed the research issue? *Answered 1 if survey or interview methodologies was appropriate for the research issue; in many cases, authors did surveys or interviews where quantitative methodologies would have been more appropriate*
6. What are the overall results of the review? *Answered 1 if the results were clearly outlined and if possible, included some numerical results*	6. Has the relationship between researcher and participants been adequately considered? *Answered 1 if this was discussed in the paper*
7. How precise are the results? *Answered 1 if the results included at least a nod to numbers, statistics or figures, since precision requires quantification*	7. Have ethical issues been taken into consideration? *Answered 1 if ethical issues were discussed*
8. Can the results be applied to the local population? *Answered 1 if the results in the paper were useful to the local population studied*	8. Was the data analysis sufficiently rigorous? *Answered 1 if there was a clearly outlined procedure for data analysis, ideally with at least a nod to numbers, statistics or figures*
9. Were all important outcomes considered? *Answered 1 if the paper included related literature from a rigorous literature search, as outlined in previous questions*	9. Is there a clear statement of findings? *Answered 1 if there was a clear summary of the findings*
10. Are the benefits worth the harms and costs? *Answered 1 if the paper was useful for the field of research this dissertation is studying*	10. How valuable is the research? *Answered 1 if the paper was useful for the field of research this dissertation is studying*

It should be noted that this analysis was performed by the PI alone, and is subject to the usual questions about bias and experience in taking into account the answers to the questions.

Also, as mentioned before in Method, the listed papers are not necessarily related to HRO and ICE, because the search methodology did not specifically seek those types of papers. It should be noted, however, that several of the papers that appeared in the search had relevance to space or aviation (e.g., Beaty et al., 2019; Ricci et al., 2012; Mogilever et al., 2018) ICE (e.g., Bell, Brown, Shanique & Mann, 2018; Håvold, Nistad, Skiri & Ødegård, 2015) and the possible HRO environment of nuclear power plants (e.g., Carroll, 2006; Chrichton & Flin, 2004; Huang & Hwang, 2009). Several were also authored or co-authored by Edmund Salas, who has studied teams working together in space exploration (e.g. Day, Gronn & Salas, 2004; Marlow, Lacerenza, Paoletti, Burke & Salas, 2004; Paris, Salas & Cannon-Bowers, 2000; Salas, Dizagranados, Klein, Burke, Stagl, Goodwin & Haplin, 2008). This does show that the subset of papers has applicability to the research question, at least so far as the topic.

Elimination by aspects showed that many papers were felled by matters such as if the authors used a systematic literature search or selected the right methodology (Q2), or if all of the relevant studies were included/if the research design was appropriate (Q3). This shows that many of the papers considered in the literature search did not make systematic criteria available for inclusion and exclusion when conducting studies, or when selecting subjects. Another common failure point was Q6: the overall results of the review, or whether the relationship between researcher and participants was adequately considered. Many literature review papers failed to clearly outline their results, while a number of qualitative studies did not outline the ethical procedures taken to ensure protection of their subjects during interviews.

Only six of the original list of 31 papers had qualitative criteria strict enough to receive positive answers for each of the 10 CASP questions, meaning they met criteria such as rigorous study designs, precisely defining inclusion and exclusion criteria, ethical procedures were considered, and the results were adequately analyzed. Of those six papers, just two studies had applicability to answering the research question for this study:"Does this paper grade or address team size with relation to error?" Admittedly, there was little attention paid to team size in either of these studies.

McEwan, Ruissen, Eys, Zumbo, Beauchamp (2017) said there was little reliability across studies to consider the impact of team size on team performance (which could include error, but not necessarily include error). The study aimed to broadly examine the concept of teamwork in a wide variety of fields, including military, aviation, health care and academic settings. The purpose of the McEwan et al. study was "to better understand the utility of teamwork training for enhancing team effectiveness", comparing those teams that had received training with those who had not, across a meta-analysis. They then examined how any teamwork interventions affected processes and/or performance across the team. The authors had planned to include team size as a moderator within the intervention, but said there was "an insufficient amount of reliable data across the studies on these variables" to conduct an analysis of this kind of subgroup. Many organizations (such as hospitals) listed the number of participants in a study, but neglected to mention the size of the participating teams, they added. Team size, as long as length of/contact time with the intervention, both had "a paucity of information available in the included manuscripts", which McEwan et al. listed as a major limitation in conducting their meta-analysis.

Similarly, Salas, Diazgranados, Klein, Burke, Stagl, Goodwin, and Halpin (2008) also examined the matter of training interventions to enhance team outcomes such as cognitive outcomes, affective outcomes, teamwork processes, and performance outcomes. They performed several meta-analytic integrations, studying a database of 93 effect sizes representing 2,650 teams. The researchers said that one mitigating factor in their study was determining the "sufficient" number of members to complete a task, because there have been mixed results examining the size of a team compared to its effectiveness; specifically, large teams are sometimes identified with better effectiveness, while sometimes they are associated with less coordination and certain process losses. They did not identify a lack of documentation concerning team size as a factor in their analysis, though. They in fact found enough studies on team size to create groupings of small teams ($n = 2$), medium teams ($2 < n < 5$) and large teams ($n \geq 5$). Further analysis of the selected studies, however, showed difficulties in isolating effects for different sizes of teams, presumably due to small sample sizes. This meant that the "diverse set of findings does not provide any appreciable level of support" for their hypothesis, which stated that team training would be most beneficial for large and small teams.

This analysis of the 31 studies thus provides insufficient information to accept either Hypothesis 1 or its null forms, and is hence, inconclusive. There is not enough information in the surveyed literature to definitively link team size with error. Error may be contained in broader terms in the literature such as "performance", but as the literature did not necessarily contain this link, it is hard to say definitively whether that is a metric of how other teams considered performance.

Hypothesis 2: Leadership and human error

Hypothesis 2 states that transformational leaders in ICE reduce the number of human errors compared with transactional leaders, based on Bass (1990). Results at the ILMAH actually showed the opposite, that transactional leadership was more effective in the ICE environment. Moreover, the approach by leaders was situational rather than fixed, as one crew member changed leadership styles between missions. While this finding might be related to the experience level of the subjects, that aspect was not considered for this study.

As explained previously, the three participating crew members at ILMAH in each of the two missions were evaluated against a Schwartz scale of values: Achievement, Benevolence, Conformity, Hedonism, Power, Security, Self-direction, Tradition and Universalism. Three coders participated in the analysis of each mission, and a Pearson correlation was performed among the coders on each occasion. In most cases, the correlation was well above 0.7. A handful of metrics had lower agreement with the coders, most particularly in instances when a Schwartz value had a low number of raw observations. For example: while analyzing Mission 2, Coder 3 found a 0 raw count for Universalism. In that case, it was impossible to perform a Pearson correlation.

The coders were instructed to apply the Schwartz values to the transcripts in the following matter: for Achievement, celebrating success of a task; for Benevolence, through crew members acknowledging each other through questions, thank-yous and the like; through Conformity, performing jokes or other actions to appear congruent with the crew; for Hedonism, sarcasm or satisfying a bodily need like burping; for Power, a crew member giving an order; for Security, speaking about matters of security or safety; for Self-Direction, a crew member making a decision or judgment (such as uttering that a task is complete); for Stimulation, describing something novel in the observations (such as planes in the sky); for Tradition, a crew member describing how something was always done; and for Universalism, referring to a shared experience such as a movie. The partial transcript below from one of the EVAs shows how one coder, who will be kept anonymous, coded the utterances:

Subject 2 – Okay. (BENEVOLENCE)

Subject 3 – There. (ACHIEVEMENT)

Subject 2 – How can I turn it on? (BENEVOLENCE)

Subject 1 – Very nice. (BENEVOLENCE)

Subject 2 – All right. Is it already recording? (BENEVOLENCE)

Subject 2 – Mmhmm … it's already recording. All right. (SELF-DIRECTION)

Pearson values for pairs of coders are compared below for Mission 1 (Table 8) and Mission 2 (Table 9). Mission 1 included the PI, Coder 1 and Coder 2. In Mission 1, while there was a range of Pearson values showing lower and higher correlations, the majority of them were well above 0.7. Mission 2 included the PI, Coder 2 and a new coder (Coder 3). The PI and Coder 2 – presumably by virtue of having worked through the analysis before – mostly scored above 0.9 correlation. Pearsons with Coder 3 scored somewhat lower, but the majority were above 0.85.

Table 8: Pearson Values using raw counts of Schwartz values, Mission 1

	PI/Coder 1	PI/Coder 2	Coder 1/Coder 2
Achievement	0.404	0.669	0.950
Benevolence	0.989	0.988	1.000
Conformity	0.984	0.797	0.677
Hedonism	0.866	0.850	1.000
Power	0.994	0.989	0.999
Security	0.933	0.721	0.922
Self-direction	0.999	0.999	1.000
Stimulation	0.500	0.990	0.371
Tradition	0.756	0.866	0.982
Universalism	-0.939	-0.661	0.879

Table 9: Pearson Values using raw counts of Schwartz values, Mission 2

	PI/Coder 2	PI/Coder 3	Coder 2/Coder 3
Achievement	0.992	0.876	0.930
Benevolence	0.957	0.963	1.000
Conformity	0.991	0.610	0.500
Hedonism	-0.870	-0.985	0.941
Power	0.999	0.998	0.995
Security	-0.099	0.423	0.860
Self-direction	0.999	0.990	0.994
Stimulation	0.327	0.941	0.629
Tradition	0.982	0.381	0.548
Universalism	-0.959	N/A	N/A

As mentioned in Methods, the transcript analysis relied on the individual coders performing raw counts of how often each subject expressed each Schwartz value during the recordings. The Pearson correlations among these Schwartz values is in Tables 7 and 8. The goal of Hypothesis 2, however, is to see how individual subjects perform in terms of leadership style. This means breaking down the Schwartz values into a form that can be converted into the leadership styles of transformative, transformational and transactional.

The first step was to obtain the percentage of time each subject expressed a particular Schwartz value during their mission. Since each coder had different values for the individual subjects they studied, an average/mean among the coders was obtained to simplify analysis. A standard deviation was computed to measure variation. The results are in Tables 10 and 11.

Table 10: Averaged Schwartz values as a function of frequency, Mission 1

	Average of PI, Coder 2 and Coder 3			
	Subject 117	**Subject 385**	**Subject 458**	**Standard Deviation**
Achievement	1.45%	2.06%	1.66%	0.25
Benevolence	37.82%	44.02%	37.34%	3.04
Conformity	1.66%	1.18%	1.76%	0.25
Hedonism	4.84%	2.39%	9.60%	2.99
Power	11.50%	13.73%	11.76%	0.99
Security	0.42%	1.18%	0.90%	0.31
Self-direction	38.61%	32.02%	30.80%	3.43
Stimulation	1.29%	0.85%	1.41%	0.24
Tradition	0.85%	0.99%	1.41%	0.23
Universalism	1.56%	1.57%	3.37%	0.85
Sum	*100%*	*100%*	*100%*	N/A

Table 11: Averaged Schwartz values as a function of frequency, Mission 2

	Average of PI, Coder 1 and Coder 2			
	Subject 1	**Subject 2**	**Subject 3**	**Standard Deviation**
Achievement	2.16%	1.27%	1.25%	0.4243426289
Benevolence	48.37%	57.20%	58.67%	4.548745859
Conformity	0.84%	1.74%	1.21%	0.36935379
Hedonism	3.86%	5.34%	6.09%	0.9265107783
Power	15.25%	10.42%	9.40%	2.551509357
Security	0.65%	0.53%	0.50%	0.06480740698
Self-direction	25.68%	20.93%	19.44%	2.660831115
Stimulation	1.02%	0.67%	1.03%	0.1673983937
Tradition	1.12%	0.63%	0.68%	0.220151463
Universalism	1.05%	1.27%	1.74%	0.2877885026
Sum	*100%*	*100%*	*100%*	N/A

ICE studies are commonly criticized for their statistical approach. A meta-analysis of ICE studies (Golden, Chang and Kozlowski, 2018) criticizes several papers for using mean, standard deviation and bi-variate correlations to measure variations in tiny populations, since these statistical measures show a wide variance. The paper notes: "Future ICE researchers should use statistical models that allow them to identify significant relationships, make causal inferences, and account for nesting", but adds no constructive suggestions for undertaking this. (*Ibid*.) Tables 9 and 10 show that benevolence, power and self-direction tend to score very high among the coded subjects, whereas security, tradition and stimulation tended to score very low.

In Table 12, Subject 117 scored first rank in Self-direction, Subject 385 scored first in Achievement, Benevolence, Power and Security, and Subject 458 scored first in Conformity, Hedonism, Stimulation, Tradition and Universalism. In Table 13, Subject 1 scored first rank in Achievement, Power, Security, Self-direction and Tradition, Subject 2 scored first in Conformity, and Subject 3 scored first in Benevolence, Hedonism, Stimulation and Universalism.

Table 12: Rankings of Schwartz Values, Mission 1

	Percentages			*Rankings*		
Subject	**117**	**385**	**458**	**117**	**385**	**458**
Achievement	1.45%	2.06%	1.66%	3	**1**	2
Benevolence	37.82%	44.02%	37.34%	2	**1**	3
Conformity	1.66%	1.18%	1.76%	2	3	**1**
Hedonism	4.84%	2.39%	9.60%	2	3	**1**
Power	11.50%	13.73%	11.76%	3	**1**	2
Security	0.42%	1.18%	0.90%	3	**1**	2
Self-direction	38.61%	32.02%	30.80%	**1**	2	3
Stimulation	1.29%	0.85%	1.41%	2	3	**1**
Tradition	0.85%	0.99%	1.41%	3	2	**1**
Universalism	1.56%	1.57%	3.37%	3	2	**1**

Table 13: Rankings of Schwartz Values, Mission 2

	Percentages			*Rankings*		
Subject	**1**	**2**	**3**	**1**	**2**	**3**
Achievement	2.16%	1.27%	1.25%	**1**	2	3
Benevolence	48.37%	57.20%	58.67%	3	2	**1**
Conformity	0.84%	1.74%	1.21%	3	**1**	2
Hedonism	3.86%	5.34%	6.09%	3	2	**1**
Power	15.25%	10.42%	9.40%	**1**	2	3
Security	0.65%	0.53%	0.50%	**1**	2	3
Self-direction	25.68%	20.93%	19.44%	**1**	2	3
Stimulation	1.02%	0.67%	1.03%	2	3	**1**
Tradition	1.12%	0.63%	0.68%	**1**	3	2
Universalism	1.05%	1.27%	1.74%	3	2	**1**

Finally, the Schwartz scale was compared to the three leadership constructs (transformational, transactional and transformative) proposed by Sarid. Sarid's leadership constructs were previously discussed in Table 4. In Tables 13 and 14 below, each individual subject is placed against transformational, transformative and transactional leadership styles.

Transformational leadership includes Achievement, Power, Self-direction and Stimulation. Transformative leadership includes Benevolence, Self-direction, Stimulation and Universalism. Transactional leadership includes Achievement, Conformity, Power, Security and Tradition. Thus, the leadership style was obtained for each subject by finding the combination of Schwartz characteristics (as characterized by the above leadership styles) with the highest rankings.

In Table 14 encompassing Mission 1, Subject 117 showed the highest ranking in transformative leadership, while Subjects 385 and 458 showed the highest ranking in transactional leadership. In Table 15 below encompassing Mission 2, Subject 3 showed the highest ranking in transformative leadership, while Subjects 1 and 2 showed the highest rankings in transactional leadership.

Table 14: Leadership styles (in gray) by ranking frequency count, Mission 1

Transformational	**Subject 117**	**Subject 385**	**Subject 458**
Achievement	3	1	2
Power	3	1	2
Self-direction	1	2	3
Stimulation	2	3	1
Frequency, #1 ranking	*1*	*2*	*1*
Frequency, #2 ranking	*1*	*1*	*2*

Transformative	**Subject 117**	**Subject 385**	**Subject 458**
Benevolence	2	1	3
Self-direction	1	2	3
Stimulation	2	3	1
Universalism	3	2	1
Frequency, #1 ranking	*1*	*1*	*2*
Frequency, #2 ranking	*2*	*2*	*0*

Transactional	**Subject 117**	**Subject 385**	**Subject 458**
Achievement	3	1	2
Conformity	2	3	1
Power	3	1	2
Security	3	1	2
Tradition	3	2	1
Frequency, #1 ranking	*0*	*3*	*2*
Frequency, #2 ranking	*1*	*1*	*3*

Table 15: Leadership styles (in gray) by ranking frequency count, Mission 2

Transformational	**Subject 1**	**Subject 2**	**Subject 3**
Achievement	1	2	3
Power	1	2	3
Self-direction	1	2	3
Stimulation	2	3	1
Frequency, #1 ranking	*3*	*0*	*1*
Frequency, #2 ranking	*1*	*3*	*0*

Transformative	**Subject 1**	**Subject 2**	**Subject 3**
Benevolence	3	2	1
Self-direction	1	2	3
Stimulation	2	3	1
Universalism	3	2	1
Frequency, #1 ranking	*1*	*0*	*3*
Frequency, #2 ranking	*1*	*3*	*0*

Transactional	**Subject 1**	**Subject 2**	**Subject 3**
Achievement	1	2	3
Conformity	3	1	2
Power	1	2	3
Security	1	2	3
Tradition	1	3	2
Frequency, #1 ranking	*4*	*1*	*0*
Frequency, #2 ranking	*0*	*3*	*2*

Results

The results showed a mix of transformative and transactional leadership styles, with no one crew member registering as transformational. A crew member participating in both missions appeared to switch leadership styles, suggesting that leadership style may change according to the context. The final step was to see which leadership style – transactional or transformative – was less prone to errors. Errors were counted for each ILMAH mission directly from the transcripts themselves, and the accounts of the crew as written in mission reports. Participants in each resolution decision were noted, as well as the person or persons who made the resolution. In Mission 1, 38 errors were recorded. Mission control fixed 8 errors alone, while crew members participated in the resolution of 30 errors. Of those 30 resolutions, 22 were from crew members predominantly showing transactional leadership, and 6 were from crew members predominantly showing transformative leadership. In Mission 2, 23 errors were recorded. Mission control resolved four errors alone, while crew members participated in the resolution of 19 errors. Of those 19 resolutions, 16 were from crew members predominantly showing transactional leadership, while 3 were from crew members predominently showing transformative leadership.

To summarize, this meant that in the majority of cases, transactional leadership was displayed during resolution of errors, such as misunderstanding procedures. While the results are counterintuitive considering the research question, there is a small body of leadership suggesting at least one aspect of transactional leadership is used in extreme environments, although the characteristics border with transformational leadership. Perhaps more research that better defines the boundary would be useful. The aspect of conscientiousness, which is needed for detail-oriented procedures, actually lies at the boundary of transactional and transformational leadership (Tejeda, Scandura, & Pillai, 2001). Specifically, conscientiousness includes both contingent rewards (an aspect of transactional leadership) and a positive attitude towards the leader (a link to transformational leadership). However, when conscientiousness is considered with other aspects within the full leadership model, at least one study found that statistical significance with transactional and transformational leadership could not be found when taking other factors into account (Hetland, Sandal & Johnsen, *ibid*). Also, other parts of the literature suggest that conscientiousness is far more appropriate for transformational leaders, since leaders who are transformational (as well servant-leadership) create work environments that allow followers to work towards goals (Lord, 1985; Keller, *ibid*), although where conscientiousness ranks in terms of positive follower behavior – and whether that is linked more to transformational leadership – is still an active field of study. (Wang, Van Iddekinge, Zhang & Bishoff, 2019).

Summary and Future Work

Validity could be determined in larger-scale ICE studies (such as in Antarctica) or it could be compared against the transcripts of the early NASA missions (e.g. Mercury, Gemini and Apollo) that are publicly available and could be used for a similar study.

The transactional leaders may do better than other leadership styles because they focus on avoiding mistakes (Bass, 1990). It is unclear from the transcripts whether the transactional leaders were focused on error avoidance ahead of the mission, because these conversations would have taken place "offline" (inside the ILMAH, when the crew members were not recorded). It is possible that the transactional leaders would have gone over the list of EVA tasks beforehand and searched for ways of following procedures and minimizing mistakes before undertaking the work (Miller, 2011; Qing, Bligh & Kohles, 2014; Morsiani, Bagnasco & Sasso, 2016), a practice that may lead to less team satisfaction due to the emphasis on fault-finding (Cummings et al., 2010; Casida & Parker, 2011; Morsiani, Bagnasco & Sasso, *ibid*). Alternatively, the transactional leaders may have been more sensitive in-situ to mistakes as they were happening, allowing for quicker resolution of any errors (Reamer, 2005), although there are worries that transactional and laissez-faire leadership are more destructive than constructive in preventing errors by this method due to not encouraging followers to seek shared goals (Schilling & Schyns, 2014).

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Summary

Hypothesis 1 stated the error rate of a team's actions is inversely related to the size of the team. Neither it or the null hypothesis can be accepted. A literature search was performed for references of team size and error in the literature, using strict parameters for the databases and the search terms. The analysis yielded 31 candidate papers, which were analyzed qualitatively using the CASP methodology. Papers were systematically eliminated by means of evaluating each question with a "yes" or "no"; after a paper received a no, it would be removed from further consideration through the elimination by aspects methodology. Of those 31 papers, only two had meaningful discussion of the team size and error variables. It is thus possible the assumptions underlying Hypothesis 1 was flawed, and that there may be no correlation between size and error. For now, though, the evidence only points to an inconclusive answer to Hypothesis 1.

Teams working in HRO and ICE must contend with the possibility of errors and ways to prevent these errors from occurring, which usually demands a multi-faceted approach examining factors such as procedures, environment, team size, human factors and much more. Human error is often cited as a crucial factor in society-critical accidents found in aviation and medicine, for example; in spaceflight, where systems are tightly coupled with each other, human error was included as one of the factors in the fatal Columbia and Challenger incidents. Determining sources of error is thus crucial not only to preventing individual incidents, but in many cases, improving societal safety as a whole.

Hypothesis 2 stated that transformational leaders in ICE reduce the number of human errors compared with transactional leaders, based on Bass (1990). Two crews in a simulated ICE environment, the ILMAH, were recorded during their EVAs and evaluated for styles of leadership, as well as types of errors and their resolutions. Transcripts of the EVAs were generated and then evaluated qualitatively by a set of three coders per mission, against a list of characteristics from Schwartz (1990) that were fed into leadership values suggested by Sarid (2016). The sample size of ILMAH participants was five people, which is similar in size to other ICE studies. The type of leadership style recorded showed a slight preference for transactional leadership, which was opposite to the hypothesis, but could be explained because the aspect of conscientiousness lies on the boundary between transactional and transformational leadership. The small sample size of participants may also have played a role in the counterintuitive results.

Unfortunately the number of studies discussing leadership approaches in ICE remains small, although approaches such as contingency-based leadership (Stuster, *ibid)* and transformational leadership (Lowe, Kroeck & Sivasubramaniam, *ibid*) were identified in larger literature reviews. Transactional leadership is not as widely discussed, which suggests that leadership approaches and their effects on error (if any) should be examined in ICE in multiple environments, including space analogs and space itself. Future missions should use a variety of methodologies, including questionnaires, surveys and listening to recordings, to obtain information about leadership style and how that affects mission operations. Once enough information is gathered, perhaps a meta-synthesis or meta-analysis of these individual missions may be performed. However, as is always the situation in ICE, it will take some time to assemble a meaningful statistical sample because each individual mission is limited to a few participants. It may be several years, or perhaps a decade, before enough reliable and rigorous information is accumulated so that researchers can make meaningful studies of leadership in ICE. It is unclear what kind of leadership is best due to the paucity of research, so the more information accumulated, the more benefit there will be for future crews.

Limitations

Missions at the ILMAH are assumed to be a representative of ICE, especially that of spaceflight. In a fashion similar to a space mission, crew members are placed into an artificial enclosed environment, they communicate with others through radio, they only go outside in spacesuits and they perform a variety of scientific and technology-focused tasks during their working hours. However, the ILMAH is not a perfect spaceflight replicator, as it does not include factors such as highly trained astronaut crews, daily work in weightlessness, a non-oxygenated environment outside of the living quarters, and the requirement to travel to and from the living quarters using a spacecraft. These factors may influence how the missions are performed and how the participants react to circumstances within the mission, such as leadership.

It is further unclear whether the ILMAH missions represent HRO, because HRO has only been proven in very limited conditions (such as crews on nuclear submarines) due to the exacting definition, and after the crews were studied for HRO characteristics over several weeks or months. (Rochlin, La Porte & Roberts, 1987). There also has been little study across the literature of HRO in relation to ICE, which indicates a possible future direction for researchers to consider when studying leadership in these environments.

Sample size in this study is limited essentially by definition of ICE, which involves subjects that are isolated and confined. The isolation restricts access to the subjects. Hence, observation is challenging and often requires proxies such as recording or journaling, which are often limited to take crew time into account. Intrusive observation is not usually consistent with achieving mission success, unless the Hawthorne effect is triggered.

The meta-synthesis set may face two challenges in generalization. First, its sources included a database that not all researchers can access – namely, the Chester Fritz Library, nominally restricted to UND scholars. The dataset was supplemented with papers found on NTRS, but the inclusion of the closed dataset may preclude wide use by other researchers. It is also probable that Chester Fritz, although it has a strong source of papers with applicability to spaceflight, does not capture all of the available literature. Meta-syntheses of other databases may thus produce very different results.

The second challenge is that of using the CASP analysis itself. While the analysis was documented here, the analysis is qualitative and relies on researcher judgment in answering the questions. Among different researchers, there may be different interpretations of the answers to the questions, which may considerably alter the results. CASP also thus not include the results of quantitative studies, which was a large factor in reducing the initial 106 studies found to the 31 that were used for later CASP analysis.

Recommendations for further research

Beyond the recommendations listed above for improving the field of research as a whole for future studies of team size, error and leadership, a selection of recommendations for future studies in a similar ICE environment are considered.

One possibility is examining transactional leadership and its possible success in error prevention by asking the ILMAH participants to keep journals describing their preparation for EVA, as well as their priorities during EVA. The journals should encompass a span of time before, during and after the mission to evaluate variation of results in isolated and non-isolated conditions. The journal entries could address how individual crew members consulted necessary procedures (for example, whether they focused on following procedures to the letter, or focused on finding possible flaws in the procedures that could lead to error-making) and then also address the approach these crew members took as errors were happening in the field – how those errors were noticed, and how they were addressed.

The Schwartz scale in relation to ILMAH participants' transcripts may be expanded to other missions and to other analogs (HERA, NEEMO, HI-SEAS and the like) to determine the efficacy of the scale and whether it has relevance for other ICE environments. In particular, it would be useful to determine the statistical validity of the scale, since in this particular instance the variance among individual coders exceeded standard deviation. A broader cross-section of participants than students from a single post-secondary institution would also be useful, to account for population bias.

Future studies could seek published literature about the degree of external reliability monitoring in larger ICE environments, to better determine how applicable these are to HRO. In support of Hypothesis 2, future studies could examine the possible link between transactional leaders and error resolution by asking crew members to keep journals of their activities before and during EVA.

ACRONYMS

EPSCoR	Experimental Program to Stimulate Competitive Research
EVA	Extra Vehicular Activity
HERA	Human Exploration Research Analog
HI-SEAS	Hawai'i Space Exploration Analog and Simulation
HRO	High Reliability Organization
ICE	Isolated, Confined Environment
IMLAH	Inflatable Lunar/Martian Analog Habitat
MDRS	Mars Desert Research Station
NASA	National Aeronautics and Space Administration
NEEMO	NASA Extreme Environment Mission Operations
NPM	New Public Management
NTRS	NASA Technical Reports Server
PI	Principal Investigator
STAMP	Systems-Theoretic Accident Model and Processes

REFERENCES

Text

Allard-Poesi, F. & Giordano, Y. (2015). Performing leadership “in-between” Earth and sky. *M@n@gement, 18*(2). 102-131.

Alnuaimi, O.A., Robert, Jr., L.P. & Maruping, L.M. (2010). Team size, dispersion and social loafing in technology-supported teams: A perspective on the theory of moral disengagement. *Journal of Management Information Systems 27*(1), 203-230.

Ark, V.A., Sipes, W., Holland, A. & Cockrell, G. (2010). *Human Behavior and Performance Support for ISS Operations and Astronaut Selections: NASA Operational Psychology for Six-Crew Operations.* Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20100036824.pdf

Avolio, B. J. & Bass, B. M. (1991). *The full range leadership development programs: Basic and advanced manuals.* Binghamton, NY: Bass, Avolio & Associates.

Bass, B.M. (1985). *Leadership Beyond Expectations*. New York: Free Press.

Bass, B.M. (1990). From transactional to transformational leadership: Learning to share the vision. *Organizational Dynamics 18*(3), 19-31.

Bass, B.M. (2000). The future of leadership in learning organizations. *The Journal of Leadership Studies 7*(3), 18-40.

Bass, B.M. & Avolio, B.J. (1990). *Transformational Leadership Development: Manual for the Multifactor Leadership Questionnaire*. Palo Alto, CA: Consulting Psychologists Press.

Berkes, H. (2016). 30 years after explosion, Challenger engineer still blames himself. *The Two-Way.* National Public Radio. Retrieved from http://www.npr.org/sections/thetwo-way/2016/01/28/464744781/30-years-after-disaster-challenger-engineer-still-blames-himself

Bierly, P.E. & Spender, J.C. (1995). Culture and high reliability organizations: The case of the nuclear submarine. *Journal of Management, 21*(4). 639-656.

Bierly, P. E., Gallagher, S. & Spender, J. C. (2008). Innovation and learning in high-reliability organizations: A case study of United States and Russian nuclear attack submarines, 1970-2000. *IEEE transactions on engineering management*, 55(3), 393-408.

Bishop, S.L., Grobler, L.C. & Schjoll, O. (2001). Relationship of psychological and physiological parameters during an Arctic ski expedition. *Acta Astronautica 49*(3-10), 261-270.

Boin, A., & Schulman, P. (2008). Assessing NASA's safety culture: The limits and possibilities of high-reliability theory. *Public Administration Review, 68*(6), 1050-1062.

Brass, S.D., Olney, G., Glimp, R., Lemaire, A. & Kingston, M. (2018). Using the patient safety huddle as a tool for high reliability. *The Joint Commission Journal on Quality and Patient Safety 44*(4), 219-226.

Brodbeck. F.C., Frese, M., Akerblom, S., Audia, G., Bakacsi, G., Bendova, H. ... & Wunderer, R. (2000). Cultural variation of leadership prototypes across 22 European countries. *Journal of Occupational and Organizational Psychology 73*, 1-29.

Brown, D.J., & Lord, R.G. (2001). Leadership and perceiver cognition: moving beyond first order constructs. In M. London (Ed.). *How People Evaluate Others in Organizations.* Mahwah, NJ: Erlbaum. 181-202.

Brown, K.M. (2004). Leadership for social justice and equity: Weaving a transformative framework and pedagogy . *Educational Administration Quarterly, 40*(1), 77-108.

Burns, J.M. (1978). *Leadership*. New York: Harper & Row.

Burns, N. (1989) Standards for qualitative research. *Nursing Science Quarterly, 2*(1), 44-52.

Burke, C.S., Shuffler, M.L., Wiese, C.W. (2018). Examining the behavioral and structural characteristics of team leadership in extreme environments. *Journal of Organizational Behavior, 39*(6). 716-730.

Burrough, B. (1998). *Dragonfly: NASA and the crisis aboard Mir*. New York: HarperCollins.

Casida J. & Parker J. (2011) Staff nurse perceptions of nurse manager leadership styles and outcomes. *Journal of Nursing Management 19*, 478-486.

Casler, J.G. (2014). Revisiting NASA as a high reliability organization. *Public Organization Review 14*(2), 229-244.

Chaudhry, A.Q. and Javed, H. (2012, April). Impact of transactional and laissez-faire leadership style on motivation. *International Journal of Business and Social Science 3*(7), 258–264.

Christianson, M.K., Sutcliffe, K.M., Miller, M.A. & Iwashyna, T.J. (2011). Becoming a high reliability organization. *Critical Care 15*(6), 314.

Ciavarelli, A. & Crowson, J. (2004). *Organizational factors in accident risk assessment.* Paper presented at Safety Across High-Consequence Industries, St. Louis, MO.

Critical Appraisal Skills Program. (2019). *About Us*. Retrieved from https://casp-uk.net/aboutus/

Cronin, M.A., Weingart, L.R. & Todorova, G. (2011). Dynamics in groups: Are we there yet? *The Academy of Management Annals 5*, 571–612.

Cummings, G.G., MacGregor, T., Davey, M., Lee, H., Wong, C.A., Lo, E. ... Stafford, E. (2010). Leadership styles and outcome patterns for the nursing workforce and work environment: a systematic review. *International Journal of Nursing Studies 47*(3), 363-385.

Dekker, Sidney. (2006). *The Field Guide To Understanding Human Error*. Hampshire, England and Burlington, VT: Ashgate Publishing.

Demchak, C.C. (1996). Tailored precision armies in fully networked battlespace: High reliability organizational dilemmas in the 'Information Age.' *Journal of Contingencies and Crisis Management, 4*(2), 93-103

Dietrich, R. & Childress, T.M. (Eds.). (2004). *Group interaction in high risk environments*. Hampshire, UK: Ashgate.

Duchesne, S.M. & Tressler, C.H. (2010). Environmental control and life support integration strategy for 6-crew operations. *40th International Conference on Environmental Systems Barcelona, Spain*. American Institute of Aeronautics and Astronautics. Retrieved from https://arc.aiaa.org/doi/abs/10.2514/6.2010-6181

Eagly, A.H., Johannesen-Schmidt, M.C. and van Engen, M.L. (2003, July). Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. *Psychological Bulletin 129*(4), 569-91.

Erwin, E.J., Brotherson, M.J., Summers, J.A. (2011). Understanding qualitative metasynthesis. *Journal of Early Intervention, 33*(3), 186-200.

Fiedler, F.E. (1967). *A theory of leadership effectiveness*. New York: McGraw-Hill.

Finlayson, K.W. & Dixon, A.. (2008). Qualitative meta-synthesis: A guide for the novice. *Nurse Researcher, 15*(2), 59-71.

Furnham, A. (1997). *The psychology of behavior at work*. London: Psychology Press.

Gatica-Perez, D. (2009). Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing 27*, 1775-1787.

Gaw, M., Rosinia, F. & Diller, T. (2018, June 1). Quality and the health system. *Anesthesiology Clinics 36*(2), 217-226.

Gawande, A. (2009). *The Checklist Manifesto: How to Get Things Right*. New York: Metropolitan Books.

Geuna, S., Brunelli, F. & Perino, M.A. (1995). Stressors, stress and stress consequences during long-duration manned space missions: a descriptive model. *Acta Astronautica 36*(6), 347-56.

Gevins, A., Smith, M., McEvoy, L. & Brown, H. (1999). Neurophysical Indices of Sustained Focused Attention: Application to Monitoring Cognitive Load, Operational Fatigue and Other Environmental Stressers. *Proceedings of the First Biennial Space Biomedical Investigators' Workshop*. 195

Golden, S.J., Chang, C-H. (D) & Kozlowski, W.J. (2018). Teams in isolated, confined and extreme (ICE) environments: Review and integration. *Journal of Organizational Behavior 39*, 701-715.

Goodwin, P. and Wright, G. (2004). How people make decisions involving multiple objectives. In *Decision Analysis for Management Judgment, 3rd Ed.* Sussex, England: John Wiley & Sons, Ltd.

Gravetter, F.J. & Wallnau, L.B. (2012) Correlation. In *Statistics for the behavioral sciences: Ninth edition*. Belmont, Calif.: Wadsworth Publishing. 509-556.

Gunderson, E.K.E. & Nelson, P.D. (1963). Adaptation of small groups to extreme environments. *Aerospace Medicine 34*, 1111-1115

Hacker, B. & Grimwood, J. (1977). Organizing Program Gemini. In *On the shoulders of Titans:*

A history of the Gemini program. Retrieved from http://history.nasa.gov/SP-4203/toc.htm.

Helmreich, R.L. and Merritt, A.C. (1998). *Culture at work in aviation and medicine: national, organizational and professional influences*. Brookfield, VT: Ashgate.

Helmreich, R.L. (2000). Culture and error in space: Implications from analog environments. *Aviation, Space, and Environmental Medicine 71*(9-11), 133-139.

Hetland, H., Sandal, G.M. & Johnsen, T.B. (2008). Followers' personality and leadership. *Journal of Leadership & Organizational Studies 14*(4), 322-331.

Heylighen, F. & Joslyn, C. (1992). What is systems theory? *Principia Cybernetica Web* (prepared for the Cambridge Dictionary of Philosophy). Retrieved from http://pespmc1.vub.ac.be/SYSTHEOR.html

Hinkin, T.R. & Schriesheim, C.A. (2008). A theoretical and empirical examination of the transactional and non-leadership dimensions of the Multifactor Leadership Questionnaire (MLQ). *The Leadership Quarterly 19*(5), 501-513.

Hoegl, M., Parboteeah, K.P. & Gemuenden, H.G. (2003). When teamwork really matters: Task innovativeness as a moderator of the teamwork-performance relationship in software development projects. *Journal of Engineering Technology Management 20*, 281-302

Hoffman, L.P. & Burrello, L.C. (2004). A case study illustration of how a critical theorist and a consummate practitioner meet on common ground. *Educational Administration Quarterly 40*(2), 268–289.

Hunt, J.G. (Summer 1999). Transformational/charismatic leadership's transformation of the field: An historical essay. *The Leadership Quarterly 10*(2), 129-144.

Jacubowski, A., Abeln, V., Vogt, T., Yi, B., Chouker, A., Fomina, E. ... & Schneider, S. (2015). The impact of long-term confinement and exercise on central and peripheral stress markers. *Physiology & Behavior 152*, 106-111.

Kanas, N., Weiss, D.S. & and Marmar, C.R. (1996). Crew member interactions during a Mir space station simulation. *Aviation, Space and Environmental Medicine 67*, 969-975.

Kanas, N. (1998). Psychosocial issues affecting crews during long-duration international space missions. *Acta Astronautica 42*(1-8), 339-361.

Kanas, N., Sandal, G., Boyd, J.E., Gushin, V.I., Manzey, D., North, R. ... & Wang, J. (2009). Psychology and culture during long-duration space missions. *Acta Astronautica 64*, 659-677.

Kanas, N., Saylor, S., Harris, M., Neylan, T., Boyd, J., Weiss, D.S. ... & Marmar, C. (2010.) High versus low crewmember autonomy in space simulation environments. *Acta Astronautica 67(7)*, 731-738.

Kanas, N. (2014). Psychosocial issues during an expedition to Mars. *Acta Astronautica 103*, 73-80

Keller, T. (1999). Images of the familiar: individual differences and implicit leadership theories. *The Leadership Quarterly 10*, 589–607.

Kelloway, E.K., Barling, J. & Helleur, J. (2000). Enhancing transformational leadership: The roles of training and feedback. *Leadership & Organization Development Journal 21*(3), 145-149.

Koopman, P.L., Den Hartog, D.N., Konrad, E., Akerblom, S., Audia, G., Bakacsi, H. ... & Wunderer, R. (1999). National culture and leadership profiles in Europe: some results from the GLOBE study. *European Journal of Work and Organizational Psychology 8*, 503–520.

Koren, M. (2018, June 22). *When a Mars Simulation Goes Wrong*. The Atlantic. Retrieved from https://www.theatlantic.com/science/archive/2018/06/mars-simulation-hi-seas-nasa-hawaii/553532/

Kozlowski, S.W.J. & Chao, G.T. (2012). The dynamics of emergence: Cognition and cohesion in work teams. *Managerial and Decision Economics 33*, 335–354.

Kraft, C. (2001). Rockets. In *Flight: My Life in Mission Control*. New York: Penguin Books. 96-112.

Lachal, J., Revah-Levy, A., Orri, M., Moro, M. R. (2017). Metasynthesis: An original method to synthesize qualitative literature in psychiatry. *Frontiers in Psychiatry, 8*, 269.

Leary, H. & Walker, A. (2018). Meta-analysis and meta-synthesis methodologies: Rigorously piecing together research. *TechTrends* 62(5), 525-534.

LaPorte, T.R. (1982). On the design and management of nearly error-free organizational control systems. In D.L. Sills, C.P. Wolf, V.B. Shelanski (Eds.). *Accident at Three Mile Island: The human dimension*. Boulder, CO: Westview Press. 185-200.

LaPorte, T.R. & Consolini, P. (1998). Theoretical and operational challenges of “high reliability organizations:” Air-traffic control and aircraft carriers. *International Journal of Public Administration, 21*(6-8), 847-852.

LaPorte, T.R. (2006). Institutional issues for continued space exploration: High-reliability systems across many operational generations – requisites for public credibility. In S.J. Dick & R.D. Launius (Eds.). *Critical Issues in the History of Spaceflight*. Washington, DC: NASA. 403-426

Lekka, C. & Sugden, C. (2011). The successes and challenges of implementing high reliability principles: A case study of a UK oil refinery. *Process Safety and Environmental Protection 89*, 443-451.

Leon, G., Sandal, G. & Larsen, E. (2011). Human performance in polar environments. *Journal of Environmental Psychology 31*, 353-360.

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science 42*(4), 237-270.

Lord, R.G., Foti, R.J. & De Vader, C.L. (December 1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational Behavior and Human Performance 34(3)*, 343-378.

Lord, R.G. (1985). Accuracy in behavioral measurement: An alternative definition based on raters’ cognitive schema and signal detection theory. *Journal of Applied Psychology 70*, 66–71.

Lord, R.G., Brown, D.J., Harvey, J.L. & Hall, R.J. (2001). Contextual constraints on prototype generation and their multilevel consequences for leadership perceptions. *The Leadership Quarterly 12*, 311-338.

Lowe, K., Kroeck, K.G. & Sivasubramaniam, N. (1996). Effectiveness correlates of transformational and transactional leadership: A meta-analytic review of the MLQ literature. *Leadership Quarterly 7*(3), 385-425.

MacNeil, R.R. & Brcic. J. Coping with the subterranean environment: A thematic content analysis of the narratives of cave explorers. *Journal of Human Performance in Extreme Environments 13*(1). DOI: 10.7771/2327-2937.1089

Maricchiolo, F., Livi, S., Bonaituo, M. & Gnisci, A. (2011). *Hand gestures and perceived influence in small group interaction*. The Spanish Journal of Psychology 14(2), 755-764.

Marks, M.A., Mathieu, J.E. & Zaccaro, S.J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review 26*, 356–376.

Martínez-Córcoles, M. (2018). High reliability leadership: A conceptual framework. *Journal of Contingencies and Crisis Management 26*(2), 237-246.

McCambridge, J., Witton, J. & Elbourne, D.R. (2014, March). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology 67*(3), 267-277.

McCurdy, H.E. (1993). *Inside NASA: high technology and organizational change in the U.S. space program.* Baltimore: Johns Hopkins University Press.

McIver D., Lengnick-Hall C., Lengnick-Hall M. & Ramachandran I. (2013). Understanding work and knowledge management from a knowledge-in-practice perspective. *Academy of Management Review 38*, 597-620

Miller, T. (2011, Sept. 11). Integrity in transactional leadership. *Advancing the Integrity of Professional Practice 135*, 35-44.

Mischel, W. (1977). The interaction of person and situation. In D. Magnusson, & D. Endler (Eds.), *Personality at the Crossroads: Current Issues in Interactional Psychology*. Hillsdale, N.J.: Erlbaum. 333-352.

Morath, J., Filipp, R. & Cull, M. (2014, October). Strategies for enhancing perioperative safety: Promoting joy and meaning in the workforce. *AORN Journal 100*(4), 376-389.

Morsiani, G., Bagnasco, A. & Sasso, L. (2016, Dec. 5). How staff nurses perceive the impact of nurse managers' leadership style in terms of job satisfaction: a mixed method study. *Journal of Nursing Management 25*(2), 119-128.

National Aeronautics and Space Administration. (2004). *An astronaut's work*. Retrieved from https://www.nasa.gov/audience/forstudents/9-12/features/F_Astronauts_Work.html.

National Aeronautics and Space Administration (2018). *About Analog Missions*. Retrieved from https://www.nasa.gov/analogs/what-are-analog-missions.

National Space Biomedical Research Institute Annual Report. (2000). *Human Performance Factors Team Program Executive Summary.* 64-65. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010002091.pdf.

Nelson, P.D. (1964). Similarities and differences among leaders and followers. *Journal of Social Psychology 63*, 161-167.

Offstein, E.H., Kniphuisen, R., Bichy, D., Childers, J. (2013). Rebuilding reliability; Strategy and coaching in a high hazard industry. *Journal of Organizational Change Management 26*(3), 529-555.

Oxford University Press. (2018). Reliability. Retrieved from https://en.oxforddictionaries.com/definition/reliability.

Padgett, J., Gossett, K., Mayer, R., Chien, W.-W. & Turner, F. (2017). Improving patient safety through high reliability organizations. (Case study). *The Qualitative Report 22*(2), 410-16.

Palinkas, L.A. (2001, June). Psychosocial issues in long-term space flight: overview. *Gravitational and space biology 14*(2), 153-163.

Palinkas, L.A. & Anderson, N.B. (2003). The psychology of isolated and confined environments. *American Psychologist 58*(5), 353-363.

Palinkas, L.A. & Suedfeld, P. (2008, January). Psychological effects of polar expeditions. *The Lancet 371*(9607), 153-163

Pearce, C.L. & Herbik, P.A. (2004). Citizenship behavior at the team level of analysis: The effects of team leadership, team commitment, perceived team support, and team size. *The Journal of Social Psychology 144*(3), 293-310.

Perrow, C. (1984). *Normal accidents: living with high risk technologies.* New York, NY: Basic Books Inc.

Prahl, A., Dexter, F., Braun, M. & Van Swol, L. (2013). Review of experimental studies in social psychology of small groups when an optimal choice exists and application to operating room management decision-making. *Anesthesia & Analgesia 117*(5), 1221-1229.

Qing, Y., Bligh, M.C. & Kohles, J.C. (2014). Absence makes the errors go longer: How leaders inhibit learning from errors. *Zeitschrift für Psychologie 222*(4), 233-245.

Rampersand, C. & Rampersand, S.E. (2007). Can medicine really learn anything from aviation? Or are patients and their disease processes too complex? *Seminars in Anesthesia, Perioperative Medicine and Pain 26*(3), 158-166.

Reamer, L.E. (2005, March). Is there an evidence-based approach to anesthesia education? *Best Practice & Research Clinical Anaesthesiology 19*(1), 137-152.

Reason, James. (2008). The Nature and Varieties of Human Error. In *The Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries*. Surrey, England: Ashgate Publishing. 29-48.

Roberts, K.H. (1990). Some characteristics of one type of high-reliability organization. *Organization Science 1*, 160-176.

Rochlin, G.I., La Porte, T.R. & Roberts, K.H. (1987). The self-designing high-reliability organization: Aircraft carrier flight operations at sea. *Naval War College Review 40*, 76-90.

Quantz, R.A., Rogers, J., & Dantley, M. (1991). Rethinking transformative leadership: Toward democratic reform of schools. *Journal of Education 173*(3), 96–118.

Salas, E., Tannenbaum, S., Kozlowski, S.W.J., Miller, C.A., Mathieu, J.E. & Vessey, W.B. (2015). Teams in Space Exploration: A New Frontier for the Science of Team Effectiveness. *Current Directions in Psychological Science, 24*(3), 200-207.

Sandal, G.M., Bergan, T., Warncke, M., Vaernes, R.J. & Ursin, H. (1996.) Psychological reactions during polar expeditions and isolation in hyperbaric chambers. *Aviation, Space and Environmental Medicine 67*, 227-234.

Sandal, G.M. Endresen, I.M., Vaernes, R. & Ursin, H. (1999). Personality and coping strategies during submarine missions. *Military Psychology 11*(4), 381.

Sandal, G.M. (2000). Coping in Antarctica: Is it possible to generalize results across setting? *Aviation, Space and Environmental Medicine 71*(9,Suppl.), A37-43.

Sandal, G.M., Leon, G.R. & Palinkas, L. (2006). Human challenges in polar and space environments. *Reviews in Environmental Science and Biotechnology 5*, 281-296.

Sandal, G.M. & Manzey, D. (2009, December). Cross-cultural issues in space operations: A survey study among ground personnel of the European Space Agency. *Acta Astronautica 65*(11-12), 1520-1529.

Schilling, J. & Schyns, B. (2014). The causes and consequences of bad leadership. *Zeitschrift für Psychologie 222*(4), 187-189.

Schwartz, S.H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology 25*, 1-65.

Schwartz, S. H. (2012). An overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture, 2*(1). https://doi.org/10.9707/2307-0919.1116

Seal, D., Bogart, L. & Ehrnhardt, A. (1998). Small group dynamics: The utility of focus group discussions as a research method. *Group Dynamics: Theory, Research and Practice 2*(4), 253-266.

Shaw, M. (1981) *Group dynamics: The psychology of small group behavior, Third Edition*. New York: McGraw-Hill Book Company.

Shields, C.M. (2010). Transformative leadership: Working for equity in diverse contexts. *Educational Administration Quarterly, 46*(4), 558-589.

Slack, K.J., Schneiderman, J.S., Leveton, L.B., Whitmire, A.M. & Picano, J.J. (2016). *Risk of adverse cognitive or behavioral conditions and psychiatric disorders: Evidence report.* National Aeronautics and Space Administration. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160004365.pdf.

Stevens, S.S. (1946, June 7). On the theory of scales of measurement. *Science 103*(2684), 677-680.

Stewart, G.L. (2006). A meta-analytic review of relationships between team design features and team performance. *Journal of Management 32*(1), 29-54.

Stuster, J. (2011). *Bold Endeavors: Lessons from Polar and Space Exploration.* Maryland: Naval Institute Press.

Suedfeld, P., Krell, R., Wiebe, R., & Steel, G.D. (1997). Coping strategies in the narratives of

Holocaust survivors. *Anxiety, Stress & Coping 10*, 153-179.

Suedfeld, P. & Steel, G.D. (2000). The environmental psychology of capsule habitats. *Annual Review of Psychology (Annual 2000)*, 227.

Suedfeld, P. (2003). Canadian Space Psychology: The Future May Be Almost Here. *Canadian Psychology 44*(2), 85-92.

Suedfeld, P. & Weiszbeck. T. (2004). The impact of outer space on inner space. A*viation, Space and Environmental Medicine 75*(7-2), C6-C9.

Suedfeld, P. (2008). *Salutongenic and pathogenic consequences of isolated, confined, and extreme environments.* Canadian Space Agency Final Report, Contract 9F007-033006/001/ ST (St-Hubert, Quebec).

Suedfeld, P., Brcic, J. & Legkaia, K. (2009). Coping with the problems of space flight: Reports from astronauts and cosmonauts. *Acta Astronautica, 65*(3-4), 312-324.

Suedfeld, P., Wilk, K.E. & Cassel, L. (2010). Flying with strangers: Postmission reflections of multinational space crews. In Vakoch, D.A. (ed.) *On orbit and beyond: Psychological Perspectives on Human Spaceflight*. Berlin: Springer Berlin Heidelberg. 185-209.

Suedfeld, P. (2010, March). Historical space psychology: Early terrestrial explorations as Mars analogues. *Planetary and Space Science 58*(4), 639-645.

Szocik, K., Abood, S. & Shelhammer, M. (2018). Psychological and biological challenges of the Mars mission viewed through the construct of the evolution of fundamental human needs. *Acta Astronautica 152*, 793-799.

Tejeda, M. J., Scandura, T. A., Pillai, R. (2001). The MLQ revisited: Psychometric properties and recommendations. *Leadership Quarterly 12,* 31-52

Tolk, J.N. & Hartley, R.S. (2010). Lessons learned from efforts to assess safety culture at the Pantex Plant. *Transactions of the American Nuclear Society103*. 493-494.

Tromp, J., Bullock, A., Steed, A., Sadagic, A. Slater, M. & Frecon, E. (1998). Small group behavior experiments in the Coven project. *IEEE Computer Graphics and Applications 18*(6), 53-63.

Turner, B. & Pidgeon, N. (1997). *Man-Made disasters.* (2nd ed.). Oxford, UK: Butterworth-

Heinemenn.

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

UND Aerospace. (2015). *Human Space Flight Laboratory*. Retrieved from http://ndspacegrant.und.edu/research/planetary-space-suit.aspx

UND Aerospace Human Spaceflight Laboratory. (2017). *2013 Experimental Program to Stimulate Competitive Research (EPSCoR) Grant.* Retrieved from http://human.space.edu/projects/planetary_exploration.htm

Vaughan, D. (1996). Preface. In *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. Chicago: University of Chicago Press. xi-1.

Wang, G., Van Iddekinge, C.H., Zhang, L., Bishoff, J. (2019, Jan.) Meta-analytic and primary investigations of the role of followers in ratings of leadership behavior in organizations. Journal of Applied Psychology 104(1), 70-106. Special Issue: Leadership.

Weiner, E.J. (2003). Secretary Paulo Freire and the democratization of power: Toward a theory of transformative leadership. *Educational Philosophy and Theory 35*(1), 89–106.

West, M.A. & Anderson, N.R. (1996). Innovation in top management teams. *Journal of Applied Psychology 6*, 680-693.

Wickens, C., Lee, J., Liu, Y. & Becker, S. (2004). Automation. In *An Introduction to Human Factors Engineering, Second Edition*. Upper Saddle River, New Jersey: Pearson Education Inc. 493-512.

Wu, R., Qianying, M., Xiong, J., Xu, Z. & Li, L. (2018). Leadership roles and group climate in isolation: A case study of 4-subject 180-day mission. *Acta Astronautica*. In press; online only at https://doi.org/10.1016/j.actaastro.2018.09.017

Yammarino, F. & Dubinsky, A. (1994). Transformational leadership theory: Using levels of analysis to determine boundary conditions. *Personal Psychology 47*(4), 787-811.

–. (1993). SpaceOps 1992: Proceedings of the Second International Symposium on Ground Data Systems for Space Mission Operations. Conference Proceedings. Doc ID. 19940019359. [PDF] Retrieved from http://hdl.handle.net/2060/19940019359.

Meta-Synthesis

Akkan, C., Erdem Külünk, M., Koçaş, C. (2016). Finding robust timetables for project presentations of student teams. *European Journal of Operational Research,* 249(2), 560-576.

Aladwani, A.M. (2002). An integrated performance model of information systems projects. *Journal of Management Information Systems*, 19(1), 185-210.

Amaya-Arias, A.C., Idarraga, D., Giraldo, V., Gómez, L.M. (2015). Effectiveness of a program for improving teamwork in Operating Rooms. *Colombian Journal of Anesthesiology*, 43(1), 68-75.

Andrews, D.R. (2011). Shooting the Moon [PDF]. Retreived from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20130014470.pdf.

Arrowsmith, J., Parker, J. (2013). The meaning of 'employee engagement' for the values and roles of the HRM function. *The International Journal of Human Resource Management,* 24(14), 2692-2712.

Aubé, C., Rousseau, V., Brunelle, E., Marques, D. (2014). The relevance of being "on the same page" to succeed as a project team: A moderated mediation model. *Motivation and Emotion* 42(6), 804-815

Beaty, D.W., Hipkin, V.J., Caudill, C.M., Hansen, R.F., Hausrath, E.M., Maggiori, C., Mccoubrey, R., Parrish, J.C., Ralston, S.J. (2019). Geological evaluation of the MSRAD field site by a human field party: Implications for rover-based exploration operations and for the future human exploration of Mars. *Planetary and Space Science*.

Bell, S.T., Fisher, D.M., Brown, S.G., Mann, K.E. (2018). An Approach for Conducting Actionable Research With Extreme Teams. *Journal of Management*, 44(7), 2740-2765.

Bodvarsson, Ö.B., Partridge, M.D. (2001). A supply and demand model of co-worker, employer and customer discrimination. *Labour Economics*, 8(3), 389-416.

Bowers, C.A., Pharmer, J.A., Salas, E. (2000). When Member Homogeneity is Needed in Work

Teams: A Meta-Analysis. *Small Group Research,* J1(3), 05-327.

Burtscher, Michael J., Levine, John M., Higgins, E. Tory (2018). On the Motivational Basis of Cue Identification in Teams. *Small Group Research*, 49(5), 519-544.

Carmel, E., Espinosa, J.A., Dubinsky, Y. (2010). "Follow the Sun" Workflow in Global Software Development. *Journal of Management Information Systems,* 27(1), 17-38.

Carroll, J.S. (Editor), (2006). Naturalistic Decision Making and Organizational Learning in Nuclear Power Plants: Negotiating Meaning Between Managers and Problem Investigation Teams. *Organization Studies*, 27(7), 1037-1057.

Castaneda, M. (2019). HIGH RISK ON THE HIGH SEAS: Marine operators are using new technology tools such as centralized shoreside risk management programs to protect remote pipelines and oil rigs from threats. *Risk Management*, 66(2), 30-35.

Chakravorty, S.S. (2010). Like riding, a bike: don't let unneeded complexity get in the way of process improvement. *Industrial Engineer*, 42(10), 47-53.

Chang, Y.H.J., Mosleh, A. (2007). Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 2: IDAC performance influencing factors model. *Reliability Engineering and System Safety*, 92(8), 1014-1040.

Choi, J.N., Sung, S. Y., Kim, M. U. (2010). How do groups react to unexpected threats? Crisis management in organizational teams. *Social Behavior and Personality: an international journal,* 38(6), 805-829.

Cohen, S.G., Ledford, G.E., Spreitzer, G.M. (1996). A Predictive Model of Self-Managing Work Team Effectiveness. *Human Relations,* 49(5), 643-676.

Cohen, S.G., Bailey, D.E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, 23(3), 239-290.

Colman, M.M., Carron, A.V. (2001). The Nature of Norms in Individual Sport Teams. *Small Group Research*, 32(2), 206-222.

Courtenay, M., Nancarrow, S., Dawson, D. (2013). Interprofessional teamwork in the trauma setting: a scoping review (Report). Human Resources for Health, 11(1).

Crichton, M.T., Flin, R. (2004). Identifying and training non-technical skills of nuclear emergency response teams. *Annals of Nuclear Energy*, 31(12), 1317-1330.

Daspit, J.J., Ramachandran, I., D'Souza, D.E. (2014). TMT shared leadership and firm performance: investigating the mediating role of absorptive capacity. *Journal of Managerial Issues*, 26(3), 219-240.

Day, D.V., Gronn, P., Salas, E. (2004). Leadership capacity in teams. *The Leadership Quarterly*, 15(6), 857-880.

Elsayed, E. A. (2000). Invited paper Perspectives and challenges for research in quality and reliability engineering. *International Journal of Production Research*, 38(9), 1953-1976.

El-Shafy, I. A., Delgado, J., Akerman, M., Bullaro, F., Christopherson, N.A.M., Prince, J. M. (2018). Closed-Loop Communication Improves Task Completion in Pediatric Trauma Resuscitation. *Journal of Surgical Education*, 75(1), 58-64.

Espinosa, J. A., Slaughter, S. A., Kraut, R. E., Herbsleb, J. D. (2007). Familiarity, complexity, and team performance in geographically distributed software development (task familiarity). *Organization Science*, 18(4), 613-631.

Filho, E. (2019). Team Dynamics Theory: Nomological network among cohesion, team mental models, coordination, and collective efficacy. *Sport Sciences for Health*, 15(1), 1-20.

Fox, J.J., Gormley, T.J. (2001). Informed maintenance for next generation reusable launch systems. *Acta Astronautica*, 48(5), 439-449.

Goll, I., Brown-Johnson, N., Rasheed, A.A. (2008). Top management team demographic characteristics, business strategy, and firm performance in the US airline industry. *Management Decision*, 46(2), 201-222.

Gong, J., Borcherding, J.D., Caldas, C.H. (2011). Effectiveness of craft time utilization in construction projects. *Construction Management and Economics,* 029(7), 737-751.

Guzzo, R.A., Dickson, M.W. (1996). Teams In Organizations: Recent Research on Performance and Effectiveness. *Annual Review of Psychology*, 47(1), 307-338.

Håvold, J. I., Nistad, S., Skiri, A., Ødegård, A. (2015). The human factor and simulator training for offshore anchor handling operators. *Safety Science*, 75, 136-145.

Hayes, T.M., Helms, M.M. (1999). Process improvement in a utility company. *Business Process*

Management Journal, 5(4), 297-310.

Hoang, C., Sabe, C., Hoang, C. (2015). Discussion. *The Journal of Thoracic and Cardiovascular Surgery*, 149(2), 505-507.

Huang, F., Hwang, S. (2009). Experimental studies of computerized procedures and team size in nuclear power plant operations. *Nuclear Engineering and Design*, 239(2), 373-380.

Hung, W. (2013). Team-based complex problem solving: a collective cognition perspective. *Educational Technology Research and Development,* 61(3), 365-384.

Jones, P. E., Roelofsma, P. H. M. P. (2000). The potential for social contextual and group biases in team decision-making: biases, conditions and psychological mechanisms. *Ergonomics*, 43(8), 1129-1152.

Jørgensen, K., Sandqvist, U., Sotamaa, O. (2017). From hobbyists to entrepreneurs: On the formation of the Nordic game industry. *Convergence: The International Journal of Research into New Media Technologies*, 23(5), 457-476.

Kidd, J.S.C., Kenneth E. (editor) (1961). A comparison of one-, two-, and three-man units under various conditions of work load. *Journal of Applied Psychology*, 45(3), 195-200.

Klimoski, R., Mohammed, S. (1994). Team mental model: construct or metaphor? J*ournal of Management*, 20(2), 403-437.

Langer, M., Braithwaite, G. R. (2016). The Development and Deployment of a Maintenance Operations Safety Survey. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 58(7), 986-1006.

Lanza, A., Simone, G., Bruno, R. (2016). Resource orchestration in the context of knowledge resources acquisition and divestment. The empirical evidence from the Italian "Serie A" football. *European Management Journal*, 34(2), 145-157.

Lawton, R., Carruthers, S., Gardner, P., Wright, J., Mceachan, R. R. C. (2012). Identifying the latent failures underpinning medication administration errors: an exploratory study. *Health Services Research*, 47(4), 1437-1460.

Lindner, M., Kalech, M., Kaminka, G. (2009). A representation for coordination fault detection in large-scale multi-agent systems. *Annals of Mathematics and Artificial Intelligence*, 56(2), 153-186.

Lutter, M. (2015). Do Women Suffer from Network Closure? The Moderating Effect of Social Capital on Gender Inequality in a Project-Based Labor Market, 1929 to 2010. *American Sociological Review*, 80(2), 329-358.

Marlow, S.L., Lacerenza, C.N., Paoletti, J., Burke, C.S., Salas, E. (2004). Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance. *Organizational Behavior and Human Decision Processes,* 144, 145-170.

Martínez-Córcoles, M., Gracia, F.J., Tomás, I., Peiró, J.M., Schöbel, M. (2013). Empowering team leadership and safety performance in nuclear power plants: A multilevel approach. *Safety Science*, 51(1), 293-301.

Mason, C.M., Griffin, M.A. (2005). Group Task Satisfaction: The Group's Shared Attitude to its Task and Work Environment. *Group & Organization Management*, 30(6), 625-652.

McCain, K.W., Salvucci, L.J. (2006). How influential is Brooks' Law? A longitudinal citation context analysis of Frederick Brooks' The Mythical Man-Month. *Journal of Information Science*, 32(3), 277-295.

McEwan, D., Ruissen, G.R., Eys, M.A., Zumbo, B.D., Beauchamp, M.R. (2017). The Effectiveness of Teamwork Training on Teamwork Behaviors and Team Performance: A Systematic Review and Meta-Analysis of Controlled Interventions. *PLoS ONE*, 12(1), p.e0169604.

Mehta, A., Feild, H., Armenakis, A., Mehta, N. (2009). Team Goal Orientation and Team Performance: The Mediating Role of Team Planning. *Journal of Management*, 35(4), 1026-1046.

Mendell, Wendell W. (Ed.) et al. (1992). *The Second Conference on Lunar Bases and Space Activities of the 21st Century, Volume 2*.

Meyer, M.H., Curley, K.F. (1995). The impact of knowledge and technology complexity on information systems development. *Expert Systems With Applications*, 8(1), 111-134.

Mogilever, N.B., Zuccarelli, L., Burles, F., Iaria, G., Strapazzon, G., Bessone, L., Coffey, E.B.J.

(2018). Expedition Cognition: A Review and Prospective of Subterranean Neuroscience With Spaceflight Applications. *Frontiers in Human Neuroscience*, Vol.12.

Morgan, B.B., Winne, P.S., Dugan, J. (1980). The Range and Consistency of Individual Differences in Continuous Work. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 22(3), 331-340.

Muller, C., Plewnia, A., Becker, S., Rundel, M., Zimmermann, L., Korner, M. (2015). Expectations and requests regarding team training interventions to promote interdisciplinary collaboration in medical rehabilitation - A qualitative study (Report). *BMC Medical Education,* 15(1).

Muscettola, N., Nayak, P.P., Pell, B., Williams, B. C. (1998). Remote Agent: to boldly go where no AI system has gone before. *Artificial Intelligence*, 103(1), 5-47.

Naikar, N. (2006). Beyond interface design: Further applications of cognitive work analysis. *International Journal of Industrial Ergonomics,* 36(5), 423-438.

Nonose, K., Kanno, T., Furuta, K. (2015). An evaluation method of team communication based on a task flow analysis. *Cognition, Technology & Work*, 17(4), 607-618.

Ornato, J.P., Peberdy, M. (2014). Applying lessons from commercial aviation safety and operations to resuscitation. *Resuscitation*, 85(2), 173-176.

Paraskevas, A. and Arendell, B. (2007). A strategic framework for terrorism prevention and mitigation in tourism destinations. *Tourism Management*, 28(6), 1560-157.

Paris, C. R., Salas, E., Cannon-Bowers, J. A. (2000). Teamwork in multi-person systems: a review and analysis. *Ergonomics*, 43(8), 1052-1075.

Park, J., Kim, J., Lee, S., Kim, J. (2018). Modeling Safety-II based on unexpected reactor trips. *Annals of Nuclear Energy,* 115, 280-293.

Pratoom, K. (2018). Differential Relationship of Person- and Task-Focused Leadership to Team Effectiveness: A Meta-Analysis of Moderators. *Human Resource Development Review*, 17(4), 393-439.

Propp, K.M., Apker, J., Zabava Ford, W.S., Wallace, N., Serbenski, M., Hofmeister, N. (2010). Meeting the Complex Needs of the Health Care Team: Identification of Nurse—Team

Communication Practices Perceived to Enhance Patient Outcomes. *Qualitative Health Research*, 20(1), 15-28. Retrieved from http://hdl.handle.net/2060/19930004784.

Ricci, M., Panos, A.L., Lincoln, J., Salerno, T.A., Warshauer, L. (2012). Is aviation a good model to study human errors in health care? *The American Journal of Surgery*, 203(6), 798-801.

Rinehart, J., Huxley, C., Robertson, D. (1994). Worker commitment and labor management relations under lean production at CAMI. *Relations Industrielles/Industrial Relations*, Fall, 49(4), 750-776.

Roberts, T. L., Cheney, P. H., Sweeney, P. D., Hightower, R. T. (2004). The Effects of Information Technology Project Complexity on Group Interaction. *Journal of Management Information Systems*, 21(3), 223-247.

Rodrigues, S., Martinho, G., Pires, A. (2016). Waste collection systems. Part B: Benchmarking indicators. Benchmarking of the Great Lisbon Area, Portugal. *Journal of Cleaner Production*, 139, 230-241.

Ryan, S., O'connor, R.V. (2013). Acquiring and sharing tacit knowledge in software development teams: An empirical study. *Information and Software Technology,* S55(9), 1614-1624.

Salas, E., Diazgranados, D., Klein, C., Burke, C.S., Stagl, K.C., Goodwin, G.F., Halpin, S.M. (2008). Does Team Training Improve Team Performance? A Meta-Analysis. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 50(6), 903-933.

Sapateiro, C., Antunes, P., Johnstone, D., Pino, J. (2017). Gathering big data for teamwork evaluation with microworlds. *Cluster Computing*, 20(2), 1637-1659.

Shearer, J., Graham, T.E., Skinner, T.L. (2016). Nutra-ergonomics: influence of nutrition on physical employment standards and the health of workers (Report). *Applied Physiology, Nutrition, and Metabolism*, 41(6 S2), S165-175.

Short, S.E., Sullivan, P., Feltz, D.L. (2005). Development and Preliminary Validation of the Collective Efficacy Questionnaire for Sports. *Measurement in Physical Education and Exercise Science*, 9(3), 181-202.

Sinclair, M.A., Siemieniuch, C.E., Haslam, R.A., Henshaw, M.J.D.C., Evans, L. (2012). The development of a tool to predict team performance. *Applied Ergonomics*, 43(1), 176-183.

Skjerve, A.B.M., Skraaning, G. (2004). The quality of human-automation cooperation in human-system interface for nuclear power plants. I*nternational Journal of Human - Computer Studies*, 61(5), 649-677.

Swezey, R.W., Owens, J.M., Bergondy, M.L., Salas, E. (1998). Task and training requirements analysis methodology (TTRAM): an analytic methodology for identifying potential training uses of simulator networks in teamwork-intensive task environments. *Ergonomics,* 41(11), 1678-1697.

Tang, T.L., Crofford, A.B. (1995). Self-managing work teams. *Employment Relations Today*, 22(4), 29-40.

Taylor, T.O., Friedman, D.J., Couture, D. (1987). Operating without supervisors: An experiment. *Organizational Dynamics*, 15(3), 26-38.

Toth, J. (1993). Manufacturing consent and resistance in peripheral production: The labor process among Egyptian Migrant Workers and Egyptian National Development. D*ialectical Anthropology*, 18(3), 291-335.

Treder, S. (2004). Baseball's New Frontier: The Expansion of 1961-62. N*INE: A Journal of Baseball History and Culture*, 2(2), 29-61.

Ulrich, B., Crider, N.M. (2017). Using teams to improve outcomes and performance (Report). *Nephrology Nursing Journal*, 44(2), 141-152.

Usman, M., Petersen, K., Börstler, J., Santos Neto, P. (2018). Developing and using checklists to improve software effort estimation: A multi-case study. *The Journal of Systems & Software*, 146, 286-309.

Van Den Bergh, J., De Bruecker, P., Beliën, J., De Boeck, L., Demeulemeester, E. A three-stage approach for aircraft line maintenance personnel rostering using MIP, discrete event simulation and DEA. *Expert Systems With Applications*.

Van Der Vegt, G.S., Bunderson, S., Kuipers, B. (2010). Why Turnover Matters in Self-Managing Work Teams: Learning, Social Integration, and Task Flexibility. *Journal of Management*, 36(5), 1168-1191.

Vera, D., Crossan, M. (2005). Improvisation and innovative performance in teams. *Organization*

Science, 16(3), 203-225.

Vera, D., Nemanich, L., Vélez-Castrillón, S., Werner, S. (2016). Knowledge-Based and Contextual Factors Associated with R&D Teams' Improvisation Capability. *Journal of Management*, 42(7), 1874-1903.

Vermuyten, H., Namorado Rosa, J., Marques, I., Beliën, J., Barbosa-Póvoa, A. (2018). Integrated staff scheduling at a medical emergency service: An optimisation approach. *Expert Systems With Applications*, 112, 62-76.

Weaver, S., Salas, E., Lyons, R., Lazzara, E., Rosen, M., Diazgranados, D., Grim, J., Augenstein, J., Birnbach, D., King, H. (2010). Simulation-based team training at the sharp end: a qualitative study of simulation-based team training design, implementation, and evaluation in healthcare.(Symposium)(Report). *Journal of Emergencies, Trauma and Shock*, 3(4), 369.

Weiss, M., Hoegl, M. (2016). Effects of relative team size on teams with innovative tasks: An understaffing theory perspective. *Organizational Psychology Review,* 6(4), 324-351.

Weiss, M., Kolbe, M., Grote, G., Spahn, D.R., Grande, B. (2018). We can do it! Inclusive leader language promotes voice behavior in multi-professional teams. *The Leadership Quarterly*, 29(3), 389-402.

Wilson, K. A., Salas, E., Priest, H. A., Andrews, D. (2007). Errors in the Heat of Battle: Taking a Closer Look at Shared Cognition Breakdowns Through Teamwork. *Human Factors: The Journal of Human Factors and Ergonomic Society*, 49(2), 243-256.

Yang, C., Yenn, T., Lin, C.J. (2010). Assessing team workload under automation based on a subjective performance measure. *Safety Science*, 248(7), 914-920.

Yingling, J.C., Goh, C., Ganguli, R. (1999). Analysis of the twisting department at Superior Cable Corporation: A case study. *European Journal of Operational Research*, 115(1), 19-35.

Zamorano, E., Stolletz, R. (2017). Branch-and-price approaches for the Multiperiod Technician Routing and Scheduling Problem. *European Journal of Operational Research,* 1257(1), 55-68.

Zheng, B., Fung, E., Fu, B., Panton, N., Swanström, L. (2015). Surgical team composition differs between laparoscopic and open procedures. *Surgical Endoscopy*, 29(8), 2260-2265.

Zhu, M., Zhang, X., Pham, H. (2015). A comparison analysis of environmental factors affecting software reliability. *The Journal of Systems & Software*, 109, 150-160.