Evidence Report:

Risk of Inadequate Critical Task Design

Anikó Sándor, Ph.D. Lockheed Martin,
NASA Johnson Space Center

Susan V. Schuh, M.S. MEI Technologies,
NASA Johnson Space Center

Brian F. Gore, Ph.D. San Jose State University,
NASA Ames Research Center

Human Research Program

Space Human Factors and Habitability

Approved for Public Release: March 12, 2013

National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas

TABLE OF CONTENTS

I. Risk of Inadequate Critical Task Design

The *Risk of Inadequate Critical Task Design* is recognized by the Human Research Program (HRP) of the National Aeronautics and Space Administration (NASA) as a risk to human health and performance in space. The HRP Program Requirements Document defines this risk. This Evidence Report provides a summary of the evidence that has been used to identify and characterize this risk.

II. Executive Summary

The Risk of Inadequate Critical Task Design relates to the definition and development of mission tasks, task flows, schedules, and procedures. In order to provide adequate task design, we need to understand relevant human capabilities and limitations for performing tasks and how these may impact workload and degrade performance on long-duration missions. We also need to understand the effect that other factors may have on human-system performance (e.g., automation).

Critical tasks can be defined as those tasks that are necessary to successfully accomplish operations and mission objectives. Task analysis is recognized as having a critical function in the design process and must be an integral part of the design phase when changes in hardware, software, and systems are easiest and less costly. Function allocation is also an important part of the design process: deciding whether a particular function will be accomplished by the human or the system, or by some combination of humans and systems. The Risk of Inadequate Critical Task Design relates to the issues arising from inappropriate definition and development of mission tasks.

Operations tempo is driven by the scheduling of mission tasks, and can affect performance, workload, and situation awareness of crewmembers. The same amount of work can be more or less taxing on crew depending on other factors such as fatigue, deconditioning, stress, and anxiety or medical conditions. Low workload levels have been associated with boredom and decreased attention to task; whereas high workload levels have been associated with increased error rates and the narrowing of attention to the possible detriment of tasks. In addition, when materials such as procedures, directions, checklists, graphic depictions, tables, charts or other published guidance are misleading or unclear, workload is further impacted, and an unsafe situation results. The severity of the consequences increases with the duration of the mission.

The Risk of Inadequate Critical Task Design includes three contributing factors, which are described in this report: a) Operational Tempo and Workload, b) Procedural Guidance, and c) Technical/Procedural Knowledge.

III. Risk Statement

Given that tasks, schedules, and procedures must accommodate human capabilities and limitations, and given that long-duration crews will experience physical and cognitive changes and increased autonomy, there is a risk that tasks, schedules, and procedures will be developed without considering the human condition, resulting in increased workload, flight and ground crew errors and inefficiencies, failed mission and program objectives, and an increase in crew injuries.

IV. Risk Overview

The Risk of Inadequate Critical Task Design relates to the definition and development of mission tasks, task flows, schedules, and procedures. In order to provide adequate task design, we need to understand relevant human capabilities and limitations for performing tasks and how these may degrade on long-duration missions. We also need to understand the effect that other factors may have on the human-system performance (e.g., the introduction of automation).

A. Designing Tasks and Procedures

Task design that maximizes system performance can be achieved when a human-centered design (HCD) process is used during the system development life cycle. HCD is characterized by early and continual user input, employed in an iterative design-test-redesign process (International Standards Organization [ISO] 13407, 1999). It focuses on making hardware/software, tasks, and related procedures usable by the human throughout a system's entire life cycle.

Lack of an HCD process often leads to inadequate task design. Inadequately designed tasks, procedures, and schedules may lead to flight and ground crew errors that threaten crew safety and compromise mission and program objectives. Designs that do not have an adequate level of usability can also lead to lack of efficiency, lack of crew satisfaction, crew frustration, and loss of productivity. Procedures that are illegible, confusing, or at the wrong level of detail negatively impact the task, putting the mission in jeopardy. Schedules that do not take into consideration human biorhythms and physical and cognitive abilities in extreme environments may affect workload that in turn may lead to poor crew performance (e.g., errors or increased task completion time). It is often the case that when tasks are not designed correctly for the human operator, additional training is needed to improve the level of performance. Inadequate task design can therefore result in additional time for training, more support during operations, redesign after deployment, and overall increased life cycle cost.

Understanding the user and the operating environment is important to ensure that design solutions meet the needs of the user within the constraints of the operating environment. This means that full awareness of the user capabilities and limitations, skills and expertise, the work environment's constraints and challenges (e.g., microgravity, isolation, small enclosed volumes), and the tasks and schedule that will be performed to accomplish the mission (e.g., piloting, maintenance, eating, sleeping) is required to ensure correctly specified task designs.

Task analysis is a method within HCD that represents tasks as a hierarchy of steps and actions that are necessary to accomplish goals. It is used to understand and document the sequence of tasks, steps, and the relationship among tasks in order to meet the needs of the user performing them. Although recognized as a critical function in design, task analysis is often erroneously overlooked until final design phases when hardware, system, and software designs are too costly to change. It is essential that task analysis be conducted as early in the design phase as possible. Task analysis should be conducted iteratively and should be frequently evaluated throughout the design and development process to allow for proper verification of crew task and system design. Furthermore, task analysis should be performed to identify the "critical" tasks, i.e., those tasks that are necessary to successfully accomplish operations and mission objectives (Kirwan and Ainsworth, 1992).

The purpose of iterative task analysis applied to spacecraft operations is to determine operator needs for established mission objectives and concepts of operations. The focus is on humans and the human-system interface and on how humans perform the task within the context of the human-system unit, rather than on the system.

Tasks are usually driven by procedures, and when written directions, checklists, graphic depictions, tables, charts, or other published guidance is inadequate, misleading or inappropriate, an unsafe situation results. Guidelines for designing task flow, schedules, and procedures are critical to ensure task and mission success.

B. Operations Concepts

An Operations Concept (OpsCon) identifies the characteristics of a system and associated tasks, which in turn integrates the perspective of the human using that system. OpsCons and the tasks that comprise them define the operational tempo by which tasks are planned and accomplished. Specific tasks that comprise an OpsCon are identified by task analysis. The definition of OpsCon requires determining which tasks are necessary to accomplish an objective, how they are expected to be performed, and who they will be performed by (Ainsworth, 2004). When these are not considered or addressed appropriately, human performance may be inadequate and may result in errors. Task analyses and OpsCons allow for the identification of task features where there is a potential for human error or other issues and problems.

C. Human Capabilities and Limitations

Consideration of the human condition as it relates to task performance is critical for addressing and optimizing performance. This includes review of tasks, schedules, training, and procedures to ensure they take into account human capabilities and limitations, as well as inevitable physical and cognitive changes that occur while on a space mission. Without consideration for these and other factors, impacts to performance can include, but are not limited to, increased workload, errors and inefficiencies, and failed mission and program objectives. The severity of the consequences increases with the duration of the mission.

Research in the area of theoretical and applied psychology identifies that humans' physical, sensory, perceptual, and cognitive capabilities have constraints that are related to performance inefficiencies, including workload increases and operator error. In the area of cognitive capabilities, for example, the amount of information that can be processed is limited by working memory (Baddeley, 1992; Miller, 1956). Therefore, information overload can be a problem when trying to accomplish tasks that load the working memory of the operator. On the other hand, information underload can lead to decreased vigilance and can lead to loss of situation awareness, i.e., being less aware of important aspects of the environment needed for the current task and future actions. For all these reasons, human capabilities and limitations should be taken into consideration in the design of tasks and associated procedures, hardware, and software.

Human capabilities and limitations can be affected greatly by the duration of a mission and the degree of subsequent deconditioning of crewmembers. Spaceflight crewmembers' strength and aerobic power of load-bearing muscles can decrease during spaceflight missions. On-orbit exercise regiments have been implemented to counteract these deficits, but to date have been only partially effective. Overall, the long-term effects of living in space and the effects on performance are still generally unknown. What is known is that perception in every modality, reaction time, motor skills, and workload can be affected while in space and thus affect performance (Legner, 2004). Thus, it is important to understand how tasks, procedures, and schedules may need to be modified as deconditioning occurs.

D. Workload

Adequate human task performance during space missions depends on usable systems, as well as acceptable workload levels (Gawron, 2000). Workload is the perceived demand on human operators to satisfy the requirements of a given task. Low workload levels have been associated with boredom and decreased attention to task, whereas high workload levels have been associated with increased error rates and the narrowing of attention to the possible detriment of tasks (Parasuraman and Riley, 1997). A crewmember given a system that is difficult to use or too many responsibilities to perform may become overloaded, and his/her performance may become degraded. The manner in which tasks are designed and in which the tasks are presented to the crew or scheduled are critical to the experience of workload. Operations tempo is driven by the scheduling of mission tasks and can affect workload and situation awareness of crewmembers. Situation awareness (SA) is the perception of environmental elements with respect to time and/or space, the comprehension of their meaning, and the projection of their status after something has changed, such as time. The same amount of work can be less or more taxing on crew depending on other factors, such as deconditioning or physiological or psychological stress. We do not know how to reliably and unobtrusively measure workload during a long-duration mission, and there are still many questions about the effects of high levels of intermittent workload or low levels of sustained workload.

E. Automation

Current spaceflight crews rely on onboard automated systems. As increasing numbers of automated systems are designed to assist the human, a synergistic relationship must be developed between the human and automation to allow the two to work together to successfully accomplish tasks. On future missions with increased flight duration and increased autonomy, crews will rely even more on these systems to provide information that is appropriate, accurate, and up-to-date. Wickens, Li, Santamaria, Sebok, and Sarter (2010) found that “routine” performance was improved, workload was reduced, and situation awareness was increased by the mere introduction of automation. However, if all tasks are automated, humans can become complacent and lose SA (Parasuraman, Molloy, and Singh, 1993; Parasuraman and Riley, 1997). In addition, increased automation will result in the need for special emphasis on task design and additional training to ensure that the crew can perform the automated tasks in the event of automation failure. Automated tasks must be carefully designed to prevent the crew from losing SA or becoming unaware of or complacent about potential hazards. These situations could ultimately result in system errors, degraded crew performance, and compromised crew and vehicle safety. A specific requirement for future long-duration missions is automated planning capabilities and tools. These tools will help autonomous crews by providing alternative plans and solutions for managing daily tasks.

The following evidence illustrates the breadth and depth of the Risk of Inadequate Critical Task Design and the factors contributing to this risk. Evidence is presented for three risk factors contributing to the Risk of Inadequate Critical Task Design: operational tempo and workload, procedural guidance, and technical/procedural knowledge. These factors are based on the Human Factors Analysis and Classification System (HFACS; Shappell and Wiegmann, 2000).

F. Dependencies & Interrelationships

The Risk of Inadequate Critical Task Design (TASK) is linked in the Risk Management and Analysis Tool (RMAT) to contributing factors from the following risks:

1. Risk of Inadequate Human-Computer Interaction (HCI)
2. Risk of Performance Errors Due to Training Deficiencies (TRAIN)
3. Risk of Performance Errors Due to Fatigue Resulting from Sleep Loss, Circadian Desynchronization, Extended Wakefulness, and Work Overload (SLEEP)

The TASK risk is concerned with tasks, schedules, and procedures. Most tasks are performed using human-computer interfaces; thus, there is a strong interaction between the TASK and HCI risks. The emphasis in TASK is on factors related to the work flow: operational tempo and workload; procedural guidance; and training for specific procedural knowledge. Because of their inherent dependence on the task design, situation awareness and usability are both considered under this risk, although there is also significant relevance to HCI. In the HCI risk, emphasis is related to the structure of information, how it is presented to the user, and the methods by which the user interacts with the information. Allocation of attention, cognitive overload, environmentally induced perceptual changes, misperception or misinterpretation of displayed data, and spatial disorientation all fall under this risk. It is particularly characterized by its emphasis on the design of displays and controls.

The TRAIN risk interacts with the TASK risk in that training affects whether a procedural knowledge will be adequate during the completion of a task. The research under TRAIN addresses best methods of training for different purposes, including individual and team activities, to obtain certain skills and knowledge.

The SLEEP risk looks at the effects of sleep deprivation and fatigue. The SLEEP risk interacts with TASK in that ground evidence indicates that fatigue may lead to performance errors and high workload, which could potentially compromise mission objectives and, consequently, the mission itself. Research areas to mitigate this risk may also include: development of a self-assessment tool for cognitive function and fatigue; light therapy for phase shifting, alertness, and mood disorders; individualized protocols for sleep-wake medication use; sleep dose-response recovery curves and individualized models for countermeasure implementation and optimal work-rest schedules; and other evidence-based means to improve individual sleep quality and reduce fatigue.

The SHFE risks also interact with risks from other HRP Elements. For example, within the Behavioral Health and Performance Element, the Risk of Adverse Behavioral Conditions and Psychiatric Disorders are contributing factors of sleep loss, work overload, cognitive impairment due to medical conditions, and operational/task-related stressors. These contributing factors impact crew workload and task performance and thus have a potential impact on the TASK risk.

G. Levels of Evidence

The levels of evidence presented in this chapter are based on the Levels of Evidence in the NASA RMAT and are as follows: Case Study, Expert Opinion, Terrestrial Data, Expert Data, and Spaceflight Incidence. Evidence presented in this chapter encompasses lessons learned from 50 years of spaceflight experience and ground-based research related to the Risk of Inadequate Critical Task Design. Portions of the evidence consist of summaries of subjective experience data, as well as non-experimental observations or comparative, correlational, and case studies. It should be noted that some evidence in this chapter is derived from the Flight Crew Integration (FCI) International Space Station (ISS) Life Sciences Crew Comments Database. Although summaries of ISS crew feedback are presented as evidence, the database is protected and not publicly available due to the sensitive nature of the raw crew data it contains.

Levels of evidence may include, but are not limited to:

- Case study
- Expert data
- Expert opinion
- Spaceflight incident data
- Terrestrial data
- Modeling

V. Evidence

The Risk of Inadequate Critical Task Design includes three core contributing factors: 1) Operational tempo and workload, 2) Procedural guidance, and 3) Technical/Procedural knowledge.

The contributing factors were derived from the Department of Defense (DoD) HFACS, the industry standard for human error categorization (DoD, 2005). The evidence about risk reduction presented in this report is organized around three types of causal risk factors, selected from the HFACS categories of error (Shappell and Wiegmann, 2000). This classification system attempts to identify the point or points in a causal chain of events that produced an accident, typically with behavior identified as an error after the fact. This approach focuses on explaining events after they happen and providing a causal chain in this explanation.

A. Contributing Factor 1: Operational Tempo and Workload

1. Operational Tempo

Operational tempo is dependent on the definition, plan, schedule, and pace of activities. ISS crewmembers currently rely on ground support teams for the majority of the planning and scheduling of daily tasks, as documented in the FCI ISS Life Sciences Crew Comments Database. Software tools, such as the Onboard Short Term Plan Viewer, provide crewmembers with detailed schedules for daily activities. Although the crew can provide input to these schedules, ground support is often relied upon to adapt and change them. The higher level of autonomy required for future long-duration missions will increase the need for automated planning capabilities and tools. Crew input and ground support, as available, will supplement the tasks generated by the automated planning tools. However, automated support for these planning tasks should allow crewmembers to manage daily tasks and ensure that these tasks are performed appropriately when ground support is unavailable. This kind of automated support will be imperative for crews whenever there will be a significant lag time in crew-to-ground communications.

Developing operational plans is a complex endeavor. Planning involves reasoning about thousands of constraints and uncertain conditions to produce a sequence of commands to be executed by humans, robots, and groups of humans and robots (Lee, 2002). Planning is optimal when all tasks to be completed are identified and unforeseen events are minimized. Future spaceflight exploration, however, must consider constraints and interactions between planned activities, as well as consideration for uncertain events in the execution environment that can create real-time impacts to mission planning (McCurdy, Connors, Pyrzak, Kanefsky, and Vera, 2006). Future surface operations will require that more resources are scheduled by fewer people and that schedules will need to accommodate increased automation and autonomy for crew activities and plans. Mission planning will require the flexibility to deal with the need for expedited plans for extravehicular activities or for commanding semi-autonomous robots.

Operational tempo is often driven by the pressure to stay on schedule irrespective of the consequences of an inappropriate decision. For instance, the Koninklijke Luchtvaart Maatschappij (KLM) / Pan Am collision at Tenerife Airport, Spain, in 1977 occurred when the pilot of one of the planes took off without having received clearance (Air Line Pilots Association [ALPA], 1977). Evidence suggests that this may have been caused by time constraints imposed by deteriorating weather, ambiguous communication between the flight crew and control tower, non-standard operating procedures, and role and responsibility issues on the flight deck.

Operational tempo may also be affected by the ability of the person to complete a task uninterrupted. In aviation, common air traffic control clearance interruptions during runway approaches may lead to increases in pilot procedure performance errors, including omission of tasks, and increases in procedure completion times (Latorella, 1996).

2. *Workload*

Workload is an important component of crewmembers' interaction with systems. Designers must consider workload when hardware and software with crew interfaces, procedures, and operations are designed. Low workload levels have been associated with boredom and decreased attention to task, whereas high workload levels have been associated with increased error rates and the narrowing of attention to the possible detriment of tasks (Sheridan, 2002).

Historically, workload has been defined in a variety of ways: as a set of test or task demands, the effort exerted to meet those demands, and the performance based on the demands (Sheridan, 2002). However, in a survey of pilots, Roscoe and Ellis (1990) found that most pilots think of workload in terms of the effort required to meet the demands of the task. In other words, it is the mental and physical effort exerted by operators to satisfy the requirements of a given task or scenario. Measuring workload is a challenge because of the need to match the workload measurement method with the evaluation objectives. As a result, measuring workload for a "system", a set of activities, groups of people, or an entire mission remains difficult. The Workload Primer (Casner and Gore, 2010) provides a description of workload measurement approaches, including their purposes, strengths, and limitations. As a summary, workload measures have been designed primarily for aviation contexts and have been applied to other highly procedural domains examining tasks over short periods of time. Operations in space, however, are characterized as highly procedural and highly repetitive tasks completed at the same time every day for extended missions lasting for more than 30 days with teams of crewmembers. This leads to questions about when and what type of repetitive procedures should be automated and the impact this has on operator performance, workload, and team collaboration, as well as which tasks will become “nuisance operations” because of their repetitive nature and their subsequent interaction with workload. Furthermore, the measure of workload for a system of activities, groups of people, and an entire mission remains as of yet difficult to put into operational terms.

Of the many concerns that a space mission faces, one of the more challenging concerns surrounds the notion of workload measurement due to the dynamic nature of the operational environment. The level of activity, the perceived difficulty of a task, the pressure of the timeframe in which a task should be performed, and the prospect of success or consequences of failure play important roles in the accuracy, efficiency, and longevity of a crew whose survival is intertwined.

3. *Measuring Workload*

Although reliable techniques exist for measuring operator workload (Hill, Iavecchia, Byers, Bittner, Zaklad and Christ, 1992), surprisingly little attention has been directed toward the question of how workload affects performance in extended missions of several months, particularly in extreme environments with expert operators, such as those encountered in space operations. Classic human factors studies have demonstrated how prolonged periods of low workload can result in boredom and loss of awareness, while periods of high workload can lead to high error rates, narrowing of attention, and decreased awareness (Sheridan, 2002). The simplest interpretation of these results might lead one to conclude that workload is best maintained at an intermediate level, in which operators are neither pressed to their limits nor left with their minds wandering.

A number of researchers have proposed the idea of designing for an optimal level of workload, in which the operator is always engaged, but never overworked (Guhe, Wenhui, Zhu, Ji, Gray and Schoelles, 2005; Hart and Wickens, 2008; Casner and Gore, 2010). Unfortunately, even this simple idea is wrought with complications. First, there is little reason to believe that any constant level of workload will avoid problems with boredom and decreased awareness. Second, studies such as that by Warm, Dember, and Hancock (1996) have challenged the simplistic notions of what tasks result in low and high workloads by demonstrating how seemingly low-workload vigilance tasks can in fact be “capacity-draining assignments” that can yield high workload measures.

Long-duration missions will possess a different performance profile from missions of short durations given their time course. Astronauts may experience long periods of low workload or outright boredom punctuated with bursts of high workload and emergency and off-nominal conditions. Maintaining a ready engagement among crews during extended periods of inactivity is essential to effective performance during periods of sudden critical activity and subsequent task design. Additionally, the importance of managing workload increases in long-duration missions in regard to crew morale because crew cohesion decreases as length of stay increases. The criteria for success of the early Moon missions, which were characterized as being relatively short, were based on the physical survival of the astronauts and the proper functioning of the vehicle - “get there and back in one piece”. As missions become longer, social cohesion will come to play a larger role in the survival of the crewmembers, the maintenance of the vehicle, and the integrity of the mission success. These issues will need to be factored into the design of tasks required by the crew. The key factor is to identify behavioral markers that are consistent with operators who are operating above their maximal or minimal workload threshold in the context of long-duration operations (Gore, Macramalla, Ahumada, and Oyung, 2010).

As summarized by Casner and Gore in the Workload Primer (Casner and Gore, 2010), four approaches to measure workload have been proposed for short-duration tasks (on the scale of hours): 1) Direct performance measures, 2) indirect measures, 3) subjective measures, and 4) physiological measures. These measures will be explained, as they provide a starting point for understanding the issues that are considered relevant in obtaining workload estimates for measuring task designs. However, these measures have not been validated for long-duration tasks, thus there is a need to investigate workload models and metrics for long-duration missions.

> (1) Direct Performance Measures of Workload
> Performance measures of workload are objective in nature and are typically directly related to operators’ performance on a task (e.g., speed, accuracy, activity rates, and task success). Performance measures of workload consider only the task being performed or the work being conducted by the human operator.
>
> (2) Indirect Measures of Workload
> Indirect measures of workload are workload estimates measured by the task performance on a secondary task being performed concurrently with a primary task. In this way, workload is estimated by measuring how much “spare capacity” the operator has. If the operator is able to perform a secondary task at the same time as the primary task, we can conclude that the primary task burdens the operator with only a low or moderate amount of workload. On the other hand, if performing a secondary task leads to a breakdown in the operator’s performance on the primary task, we can conclude that the primary task absorbs most of the operator’s resources and that the operator is nearing the peak of their capacity to do work.

(3) Subjective Measures of Workload
Subjective workload measures ask the human operator to describe the workload they experience when performing a task. Subjective workload measures do not attempt to measure anything about the task that the user is performing or the user's performance on any task. Subjective workload measures focus entirely on the human operator's feelings about the workload associated with completing a given task in a specific situation. Subjective measures can include measurement techniques such as the NASA Task Load Index and the Bedford workload scale.

(4) Physiological Measures of Workload
Physiological measures of workload attempt to associate physiological changes with levels of workload. For a time, researchers hoped that physiological measures could be found that represented a truly objective workload measurement (i.e., one that did not rely on assumptions about how people perceive workload or on subjective ratings provided by human operators). Although many physiological measures have been investigated (e.g., heart rate and variability and evoked potentials), no one measure has proven to definitively capture the construct of workload.

4. *Workload-Related Incidents*

High cognitive workload has been cited as a causal reason for several aviation accidents. Wiegmann et al. (2005) analyzed accident data from general aviation and found skill-based errors were associated with nearly 80% of these accidents. Skill-based errors included inappropriate handing of aircraft and inattention to flight instruments or checklist procedures, all of which could be attributed to high workload.

On American Airlines Flight 965, a failure in automation during a critical phase of flight resulted in high pilot workload and, ultimately, a crash killing 159 passengers (Ladkin, 1996). In 1976, a mid-air collision occurred between British Airways Flight 476 and Inex-Adria Aviopromet Flight 550. All passengers and crew on both planes were killed. The air traffic controller responsible for directing the planes was overloaded with information in a congested airspace. The controller was working alone without an assistant controller. Upon realizing the collision course (which would have actually been a near-miss), the air traffic controller started talking in his native Croatian language, likely due to extreme stress, and did not realize the British Airways pilot could not understand. The controller accidently directed the planes into a collision course (Air Accidents Investigation Branch [AAIB], 1977).

A NASA example of workload as a risk factor comes from the collision in June 1997 between the Russian spacecraft Progress 234 and the Mir Space Station, which caused the pressure hull to rupture and nearly led to the Mir being abandoned. High workload and stress of the crew due to repeated system failures likely contributed to reduced vigilance capabilities (Ellis, 2000).

As illustrated in the above examples, operator workload is often the culprit behind many accidents and incidents in complex operational environments. A challenge arises when we use these past events and infer causality from them. Consequently, computational methodologies have been proposed as a cooperative approach to be used with human-in-the-loop simulations to better understand the manner in which the human and system interact (Gore, 2002; Hooey and Foyle, 2008).

Behavioral markers are needed for predicting human performance. A prototype mobile screening system that is being developed as part of the Future Attribute Screening Technology (FAST) project for the Department of Homeland Security is a candidate approach that could be leveraged and applied in the space context. The prototype screening system monitors individuals' stress through an evaluation of behavioral markers (Martin et al., 2010). Stress and workload often occur together (Hancock and Szalma, 2008). The FAST project integrates commercial off-the-shelf sensors to measure physiological, behavioral, and auditory information in order to predict passenger stress levels. This mobile screening system measures pulse rate, skin temperature, breathing, facial expression, body movement, pupil dilation, and other "psycho physiological / behavioral patterns" to stop "unknown terrorists". Developing such a mobile screening system could provide real-time information to the crew so that they can reallocate their tasks to optimize their performance.

Analysis of the data from the recently completed Mars 500 Study on long-duration isolation needs to be analyzed to determine suitability for populating candidate measures of workload for long-duration operations, as well as to determine whether common factors can be identified as drivers of operator workload during long-duration operations. Once a review of this study data has been completed, a more detailed understanding of the daily activities can be obtained and used to scope the critical variables that appear to drive workload in the long-duration time period. The Workload Directed Research Project identified a number of causal factors by analyzing available literature and crew debrief reports and developed two conceptual models of the underlying drivers of workload. One of the conceptual models views workload as related to the four 'more' common drivers/factors that underlie operator workload as populated from Crew Reports, and the second conceptual model relates workload to a workload-dose, a model that relates time criticality, dosage, and task stressfulness to each other (Gore, Macramalla, Ahumada, and Oyung, 2010). The Mars 500 Study can be used to flesh out the conceptual model of long-duration workload.

Another candidate factor that should be considered includes the impact that delays between training and actual task performance will have on operator workload and subsequent performance. It is expected that with increased delays since training, given that the duration of the mission is going to get longer, the crew will face increased workload in accessing the required knowledge for their tasks. From commercial aviation, inadequate crew knowledge of automated systems was a factor in more than 40% of accidents and 30% of serious incidents between 2001 and 2009 (Abbott, 2010). Abbott argued that flight crews are not properly trained for modern cockpits and that there is a need for radical change in their recurrent training and in the standard operating procedures of airlines. Although this risk is outlined in the context of commercial aviation, it is highly relevant for long-duration space operations.

Research has shown that automation can reduce operator workload and fatigue. Furthermore, when automation is introduced, the nature of workload shifts from motor (e.g., manual control) to cognitive (e.g., supervisory tasks such as monitoring, projecting, and analyzing) (Parasuraman and Hancock, 2007; Sheridan and Parasuraman, 2000; Billings, 1997). When patterns of workload shift in such ways, new system vulnerabilities arise (National Research Council [NRC], 1993). These vulnerabilities include decreased skill and the human-out-of-the-loop phenomenon (Sarter and Woods, 1997; Sarter and Woods, 1995). As the automation becomes more pervasive, such vulnerabilities may contribute to safety risks. Computational approaches and tools are currently being developed to assist space mission planners and remotely piloted vehicle system designers in selecting function allocation strategies based on a stages and levels model of human-automation interaction (Sebok, Wickens, Gacy, Behon, Scott-Nash, Sarter, Li, Gore, and Hooey, 2012; Gacy, Wickens, Sebok, Gore, and Hooey, 2011).

5. *Workload and Automation Interaction*

When workload exceeds the operational capabilities of the human, automation will need to be developed to support the crew during long-duration mission operations. The main risk is that no basic process models exist to predict the impact that long durations have on human workload and the subsequent impact this workload effect will have on performance.

Human performance model renditions of the tasks and procedural design sets that are representative of long-duration mission operations, including critical event responses, are non-existent. As such, there is a need to develop re-usable procedural task sets for baseline long-duration mission operations that can then be tweaked to represent events requiring critical responses required by the crew.

6. *Task Design and Information Feedback Interaction*

Behavioral change is unlikely unless the operators are provided with accurate feedback on their performance. There is a great need to know when workload is approaching a given threshold, at which point performance will be likely to decrease. Feedback, or knowledge of results, will allow the rational human operator to make changes to the manner in which they intend to complete the task set, whether it be replacing a task with a shortcut, removing a task entirely, or rescheduling the task so that it can be completed at a later time. Computational representations of times when workload will be reaching/nearing a threshold can be evaluated in simulations prior to mission operations, while real-time tracking of workload trend information can be provided to the crew in situ.

B. Contributing Factor 2: Procedural Guidance

Procedural guidance refers to the structure, content, and presentation of procedures. Procedural guidance is a contributing factor to inadequate task design when written direction, checklists, graphic depictions, tables, charts, or other published guidance is inadequate, misleading, or inappropriate, which creates an unsafe situation.

Spaceflight crewmembers are required to obtain, process, and maintain complex and large amounts of information to execute spaceflight missions. This information must be clearly presented. If information is not presented clearly, the user may process displayed messages incorrectly and misinterpret, overlook, or ignore the original intent of the information. Task performance can be jeopardized and mission success can be put at risk when procedural information is not processed as intended.

Procedures provide step-by-step instructions for completing a task. Procedures inform the user on the sequential order of tasks and often provide feedback on the outcome of the task (Salas, Wilson, Priest, and Guthrie, 2006). Operator error becomes more likely when procedures are written incorrectly or when procedures are overly complicated. Critical information can be missed, which leads to a situation where the procedures are inadequate for the user to complete the task at hand. According to the ISS Crew Comments Database, while on-orbit, crewmembers have been using paper checklists to assist them with their required tasks for many years. However, navigating through paper checklists has been difficult. Procedures are coded with specialized symbols, abbreviations, boundary delimiters, and spatial configurations that collectively require extensive training to decipher and understand. Individual instructions in these checklists frequently take the form of conditional expressions (IF-THEN-ELSE statements), which the crewmembers must evaluate by manually crosschecking systems or flight status information on cockpit instruments and displays. The outcome of the evaluation of the logical expression determines which path should be taken through the remainder of the checklist, and that path, in turn, determines which subsequent instructions have to be carried out and in what order. Wrong choices based on inappropriate assessment of the state of information presented can cause a risk to crew safety (Hudy and Woolford, 2008).

The amount of information that crews need to assimilate and use on a daily basis is extensive (Figure 1). For instance, ISS crews are loaded with information regarding operations and procedures for tasks. Some ISS electronic procedures and formats have been especially difficult to use. Frequently, crewmembers spend excessive amounts of time navigating between various menus because the procedures are difficult and lengthy, or occasionally because they contain unnecessary information. As documented in the FCI ISS Life Sciences Crew Comments Database, the structure and content of procedures have contributed to inadvertently skipping steps in the procedure that then resulted in poor task execution. In general, procedures are thought to be too detailed, especially for simple operations. Pictures and diagrams, considered helpful for many procedures, are not always integrated appropriately. In addition, some procedures reference multiple steps from other procedures. Locating the necessary steps costs the crew additional time and has resulted in missed or skipped steps. Overall, usability of procedures has been an ongoing issue for the ISS and emphasizes the need for common human factors standards and simplification where possible in procedure development (Baggerman, Rando, and Duvall, 2004).

Figure 1 Photo ISS015-E-17702 Courtesy of NASA — Cosmonaut Fyodor N. Yurchikhin, Expedition 15 commander representing the Russian Federal Space Agency, holds a camera while looking over procedure checklists in the Zvezda service module of the ISS (NASA Human Spaceflight Gallery, 2008)

Further issues associated with procedures that have occurred during ISS missions are directly related to the provision of too much information, lack of diagrams and schematics to illustrate necessary information, and confusion and missed steps caused by multiple links in procedures. These issues have caused frustration among crewmembers and have directly affected efficient task performance because the information needed for a given task has not always been presented in a usable format (Rando et al., 2005). Procedures can often complicate or impede the performance of daily tasks. Procedures may call for an inadequate number of crewmembers to perform a task, or the specified duration for a task may be inappropriate. Progress is being made on improving procedures and enhancing crewmembers' abilities to acquire information by including more graphic content (e.g., diagrams and images). The goal is to improve the procedures so that they better reflect how operations are actually conducted.

Methods for designing procedures are often based on the collective knowledge and experience of system designers, simulator trainers, and astronauts to create what is considered the "best" way to mitigate and recover from an abnormal or emergency situation. It was realized that this current method was difficult to use for future exploration procedures due to the increased reliance on automation and the complexity of the systems themselves that are certain to be required for future space missions. Specifically, procedure writers do not understand the integrated nature of the system and its components and often do not verify that the procedures that call these integrated systems are correct. In addition, there is no assurance that the selected action sequence is indeed the "best" with regard to execution time, mitigation of failure, and maximization of recovery. Based on this, a formal approach for the design of emergency procedures and recovery sequences was proposed by Degani and colleagues (2005). The approach uses mathematical methods and tools to analyze, verify, and synthesize action sequences, which can be used to verify the correctness, suitability, efficiency, reliability, and safety of an existing or proposed procedure. This approach has been used successfully in the commercial aviation domain and has been incorporated in a new standard for cockpit procedures dealing with in-flight fires, and it can be applied to all exploration systems to improve training and procedures (Heymann, Degani, and Barshi, 2007).

Desaulniers, Gillan, and Rudisill (1988) conducted two experiments to investigate the effect of text formats on operators' performance using computer-based procedure displays. In Experiment 1, procedures were presented in text, extended-text, and flowchart formats. Text and extended-text were structured prose formats differing in the spatial density of presentation. The flowchart format differed from the text format in both syntax and spatial representation. Subjects were asked to use the procedures to diagnose a hypothetical system anomaly. The results indicated that accuracy was higher with the flowchart format. Although overall task completion times did not differ across formats, the flowchart format required significantly less time for step implementation. Furthermore, the follow-on study showed that completion times for flowchart procedures decreased with increasing window size; however, accuracy of performance decreased substantially. These studies show that the presentation of procedures has a strong effect on performance.

Crew performance of tasks on the ISS, such as extravehicular activity, maintenance, and medical tasks, relies heavily on the provision of adequate procedures and ultimately a strongly defined operational concept (Rando et al., 2005). Comments reported during ISS post-increment reviews have identified that poor design of ISS procedures has impeded crew task performance by preventing the completion of scheduled activities within the allotted time. Well-designed procedures play a critical role in ensuring optimal on-schedule crew task performance; inadequately structured procedures will lead to a reduction in human task performance.

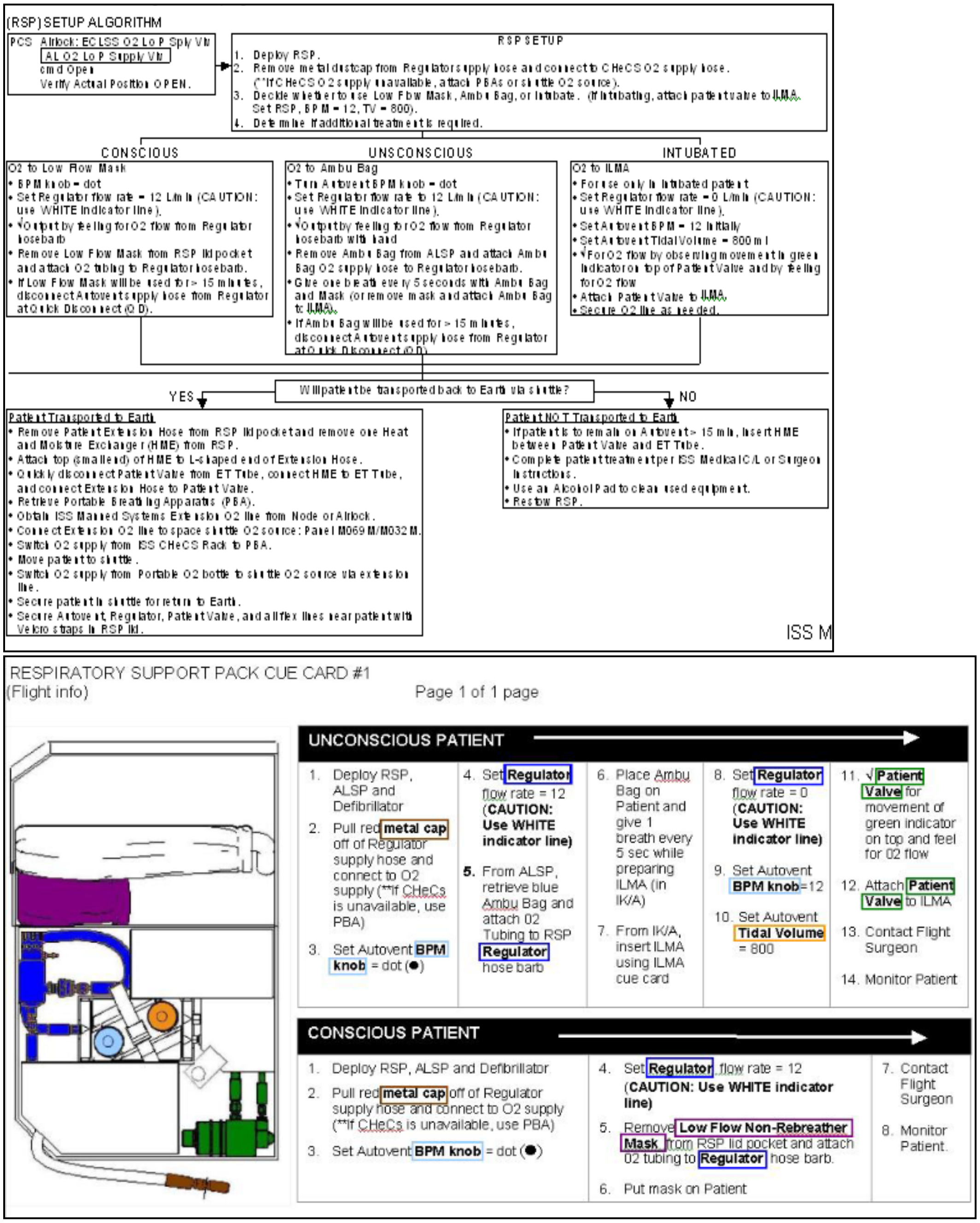
Defining tasks without an understanding of the task environment also leads to inadequate task design. Therefore, analyzing and understanding realistic operations concepts is crucial to providing procedures that work effectively and support safety and efficiency. Loukopoulos, Dismukes, and Barshi (2009) conducted a study that significantly reduced pilot errors by a systematic revision of aircraft cockpit procedures. Procedure revisions were initiated with a task analysis viewpoint, which focused on identifying concurrent tasks even though they had been designed as if they were to be accomplished in isolation. The authors focused on multitasking, i.e., handling more than one task at the same time. The approach analyzed the sequence of tasks performed by the captain and first officer, from preflight through all flight phases, including power-down. They found that procedures for these tasks are typically written for the crew only, omitting the complexities introduced by interactions with ground support and the flight crew. This research showed that the operating environment for the procedures was linear, predictable, and controllable. In contrast, the real environment included a variety of disturbances requiring changes to the task sequence. Many of the disturbances were unpredictable or uncontrollable. The researchers also identified a large number of crew omissions. Four distinct patterns were found that frequently resulted in these types of errors: interruptions and distractions, tasks that cannot be executed in the planned sequence, unanticipated new tasks, and multiple tasks that must be interleaved. After completing the analysis, the authors supported one airline's program to review and revise its procedures. The final product was implemented and resulted in an error rate reduction of over 80%. In addition, the revised procedures were learned more easily, improving training efficiency. This human error research, within the context of real-world operations, illustrates the importance of designing procedures that consider the complete system.

Real-time fault management during ascent and entry ranks among the most safety-critical of spacecraft operations. Consequently, crew performance while conducting these tasks must be supported by the highest possible quality fault management procedures and interfaces, as well as the most effective human-automation function allocation choices, with a minimum of display real estate. Two operational concepts for fault management procedures were compared by the studies conducted at the NASA Ames Research Center by Hayasi and colleagues (2007). The studies resulted in the following guidelines for procedures:

- Wherever possible, minimize operators' need to process cluttered, text-rich display formats by automating fault-management task elements that require text processing. If automating those elements is not possible, text-based displays should be structured to reduce searching requirements.
- Save display real estate by displaying just the current focus line(s) plus the following three lines (the next to-be-completed step in the checklist).

- Allow operators to be able to view the electronic procedure viewer and the relevant system summary display format (the display format containing the “soft” system mode reconfiguration interfaces) simultaneously.

Performance degradation due to inadequate ISS task design was illustrated during a ground-based study to test the usability of the procedure (as written on a “cue card,” Figure 2) for the Respiratory Support Pack (RSP), a piece of ISS medical equipment, to support redesign of the cue card (Hudy, Byrne, Smith, and Whitmore, 2005). The RSP was designed for use during medical contingencies involving respiratory distress, and it was therefore expected that the complicated RSP cue card procedure would be used in time-critical situations when a crewmember’s life could depend on the outcome. During the study, data were collected as subjects executed the procedure checklists. Results demonstrated that some procedures, equipment design, and labeling could be a source of errors and ultimately a risk to crew health. The procedures and the sequence of equipment use did not enable a crewmember to establish a patient’s airway in the time necessary to prevent irreversible brain damage. The Crew Medical Officer typically receives very limited training in using the medical equipment, and the cue cards thus hold vital information on how to execute the procedures. This example illustrates the importance of appropriate procedures and support for training to ensure that tasks can successfully be performed, especially in the case of an emergency. The cue card was subsequently redesigned to support the task (Figure 2).



(RSP) SETUP ALGORITHM

PCS Airlock: ECLSS O2 Lo P Sply Vlv
AL O2 Lo P Supply Vlv
cmd Open
Verify Actual Position OPEN.

RSP SETUP
1. Deploy RSP.
2. Remove metal dustcap from Regulator supply hose and connect to CHeCS O2 supply hose.
(**If CHeCS O2 supply unavailable, attach PBAs or shuttle O2 source).
3. Decide whether to use Low Flow Mask, Ambu Bag, or Intubate. (If intubating, attach patient valve to ILMA. Set RSP, BPM = 12, TV = 800).
4. Determine if additional treatment is required.

CONSCIOUS
O2 to Low Flow Mask
- BPM knob = dot
- Set Regulator flow rate = 12 L/min (CAUTION: use WHITE indicator line).
- √Output by feeling for O2 flow from Regulator hosebarb
- Remove Low Flow Mask from RSP lid pocket and attach O2 tubing to Regulator hosebarb.
- If Low Flow Mask will be used for > 15 minutes, disconnect Autovent supply hose from Regulator at Quick Disconnect (QD).

UNCONSCIOUS
O2 to Ambu Bag
- Turn Autovent BPM knob = dot
- Set Regulator flow rate to 12 L/min (CAUTION: use WHITE indicator line).
- √Output by feeling for O2 flow from Regulator hosebarb with hand
- Remove Ambu Bag from ALSP and attach Ambu Bag O2 supply hose to Regulator hosebarb.
- Give one breath every 5 seconds with Ambu Bag and Mask (or remove mask and attach Ambu Bag to ILMA).
- If Ambu Bag will be used for > 15 minutes, disconnect Autovent supply hose from Regulator at Quick Disconnect (QD).

INTUBATED
O2 to ILMA
- For use only in intubated patient
- Set Regulator flow rate = 0 L/min (CAUTION: use WHITE indicator line).
- Set Autovent BPM = 12 initially
- Set Autovent Tidal Volume = 800 ml
- √For O2 flow by observing movement in green indicator on top of Patient Valve and by feeling for O2 flow
- Attach Patient Valve to ILMA.
- Secure O2 line as needed.

Will patient be transported back to Earth via shuttle?

YES

Patient Transported to Earth
- Remove Patient Extension Hose from RSP lid pocket and remove one Heat and Moisture Exchanger (HME) from RSP.
- Attach top (small end) of HME to L-shaped end of Extension Hose.
- Quickly disconnect Patient Valve from ET Tube, connect HME to ET Tube, and connect Extension Hose to Patient Valve.
- Retrieve Portable Breathing Apparatus (PBA).
- Obtain ISS Manned Systems Extension O2 line from Node or Airlock.
- Connect Extension O2 line to space shuttle O2 source: Panel M069M/M032M.
- Switch O2 supply from ISS CHeCS Rack to PBA.
- Move patient to shuttle.
- Switch O2 supply from Portable O2 bottle to shuttle O2 source via extension line.
- Secure patient in shuttle for return to Earth.
- Secure Autovent, Regulator, Patient Valve, and all flex lines near patient with Velcro straps in RSP lid.

NO

Patient NOT Transported to Earth
- If patient is to remain on Autovent > 15 min, insert HME between Patient Valve and ET Tube.
- Complete patient treatment per ISS Medical C/L or Surgeon instructions.
- Use an Alcohol Pad to clean used equipment.
- Restow RSP.

ISS M

RESPIRATORY SUPPORT PACK CUE CARD #1
(Flight info)
Page 1 of 1 page

UNCONSCIOUS PATIENT

1. Deploy RSP, ALSP and Defibrillator
2. Pull red **metal cap** off of Regulator supply hose and connect to O2 supply (**If CHeCs is unavailable, use PBA)
3. Set Autovent **BPM knob** = dot (●)
4. Set **Regulator** flow rate = 12 (**CAUTION: Use WHITE indicator line)**
5. From ALSP, retrieve blue Ambu Bag and attach 02 Tubing to RSP **Regulator** hose barb
6. Place Ambu Bag on Patient and give 1 breath every 5 sec while preparing ILMA (in IK/A)
7. From IK/A, insert ILMA using ILMA cue card
8. Set **Regulator** flow rate = 0 (**CAUTION: Use WHITE indicator line)**
9. Set Autovent **BPM knob**=12
10. Set Autovent **Tidal Volume** = 800
11. √ **Patient Valve** for movement of green indicator on top and feel for 02 flow
12. Attach **Patient Valve** to ILMA
13. Contact Flight Surgeon
14. Monitor Patient

CONSCIOUS PATIENT

1. Deploy RSP, ALSP and Defibrillator
2. Pull red **metal cap** off of Regulator supply hose and connect to O2 supply (**If CHeCs is unavailable, use PBA)
3. Set Autovent **BPM knob** = dot (●)
4. Set **Regulator** flow rate = 12 (**CAUTION: Use WHITE indicator line)**
5. Remove **Low Flow Non-Rebreather Mask** from RSP lid pocket and attach 02 tubing to **Regulator** hose barb.
6. Put mask on Patient
7. Contact Flight Surgeon
8. Monitor Patient.

Figure 2 RSP information card before evaluation and after evaluation and redesign

Issues related to usability and human capabilities can also arise when safety precautions are factored into design for on-orbit operations, but not for preflight ground activities. On January 27, 1967, when the Apollo 1 crew initiated what should have been a nominal countdown drill, disaster struck and fire erupted in the command module at Kennedy Space Center's Pad 34. All three crewmembers lost their lives because they were unable to open the hatch and escape the command module to safety (Kranz, 2000). Many factors contributed to this incident, including the inadequate hatch design. The hatch was designed to open inward, which was impossible for a human to open at the pressure levels within the vehicle while it was on the ground. Procedures or processes were not in place to deal with this type of emergency event. Therefore, crewmembers did not have the necessary procedural or technical guidance to handle an event of this magnitude, which led to the worst possible consequence: loss of life. Little consideration had been given to risks and hazards associated with any of the preflight activities, only those pertaining to spaceflight, and this scenario contributed to the unsafe situation and ultimately the crewmembers' deaths.

There are many examples from airplane incidents that illustrate the importance of good procedure design. Northwest Airlines Flight 255 in 1987 (National Transportation Safety Board [NTSB], 1987) crashed shortly after takeoff, killing all but one passenger. As the plane lifted off the runway, it began to roll from side to side, went into a stall, rolled into a pole at the end of the runway, and then rolled onto a road, hitting vehicles and bursting into flames when it hit an overpass. The National Transportation Safety Board concluded that the probable cause of the accident was the failure of the pilots to use the taxi checklist, such that the aircraft flaps and slats were retracted and thus not properly configured for takeoff.

Methods of information presentation can play a large role in the retention of procedural and technical data and determine operations success or failure. Under Human Research Program (HRP) funding, a study was conducted to collect feasibility data for two prototype display types for viewing real-time medical procedures and just-in-time (JIT) training. A hand-held Personal Digital Assistant (PDA) and a Head-Mounted Display (HMD) system were compared while performing a simulated JIT medical procedure using ISS flight-like equipment (Figure 3). Participants who performed a simulated medical procedure with the HMD system and the PDA reported that on both devices, too little information could be shown to aid situation awareness of how far into the procedure they were.

Figure 3 Participants perform simulated JIT medical procedures

Another study addressed information presentation methodologies by focusing specifically on assessing multiple versions of checklist information. The aim of this study was to test for differences in performance and user opinion of the current NASA *International Space Station Integrated Medical Group (IMG) Medical Checklist* (JSC-48522), as well as two alternatively organized versions. Each version contained the organization of the emergency and non-emergency medical procedures and diagnoses. The original version featured both alphabetical and anatomical organization of the medical procedures. The two alternate versions offered an anatomical organization and an alphabetical organization.

The results showed that the anatomical and alphabetical versions led to shorter task completion times than the checklist version.

C. Contributing Factor 3: Technical/Procedural Knowledge

Technical and procedural knowledge becomes a contributing factor to inadequate task design when an individual has been adequately exposed to the information needed to perform the mission element but did not absorb it. Lack of knowledge does not imply deficiency in the training program in this case, but rather the failure of the individual to absorb or retain the information. This lack of absorption or retention of technical or procedural knowledge may be due to multiple factors; for example, training on a specific activity may have been presented too early prior to the task. For example, if crewmembers are trained on procedures that do not seem immediately relevant, although they will understand the information, they may not retain it adequately. On the other hand, if the training is provided as just-in-time training when the knowledge is relevant to an upcoming task, crewmembers will absorb the information.

Spaceflight crewmembers are required to obtain, process, and maintain information in increasing quantity and complexity to execute spaceflight missions. Information must be easily recognized and understood. If information is not presented clearly, the user may process the message incorrectly and misinterpret, overlook, or ignore the original intent of the information. When information is not processed as intended, task performance can be jeopardized and mission success can be put at risk.

During their pre-mission training flows, spaceflight crewmembers are introduced to procedures and large amounts of supplementary information that may be needed at different stages of the mission. While scheduling constraints can arise throughout a training flow, introductions to procedures and the provision of technical knowledge should be provided at the appropriate time both pre-flight and while on-orbit, based on the crew's schedule, to encourage knowledge retention and recall. Various methodologies for the presentation of procedures should be considered to accommodate user preference and operations constraints, such as the performance of a task inside of a science glove box within a space vehicle. All of these considerations for procedures can assist in proper absorption and retention of procedural and technical guidance to optimize task performance and guarantee safe operations.

The primary modes for providing spaceflight mission and operations information to the crew are through procedures and training. However, information can be conveyed through other means as well, including crew-to-crew and crew-to-ground communication, robotics, and automation. Regardless of the mode used to acquire information, it is important that the process be matched with human capabilities and limitations. This will ensure that the right information acquisition methods are being used, the duration of the acquisition process is appropriate for the type of information being transferred, the length of time after learning is adequate enough to access the information when needed, adequate metrics for monitoring the methods are in place, and just-in-time training is effective and available when needed. Proper definition and execution of these elements will ensure successful acquisition of information in the space environment, thus leading to mission success.

Training for tasks that will be performed under extreme conditions and environments can lead to less than adequate procedural knowledge. This can be due to the fact that simulated environments and ground-based, full-scale models or mockups cannot be completely representative of flight conditions. Representing a true 0g environment on the ground has presented many challenges for training and information presentation; thus, simulations may provide inadequate information during preflight training. Stowage is one example demonstrating where 1g training provides an inadequate representation of information needed for task execution in 0g. As documented in the FCI ISS Life Sciences Crew Comments Database, a true representation of the stowage of equipment and materials onboard the ISS is very difficult to achieve on the ground and can create issues for the crew. Stowage mockups in 1g are limited because gravity restricts operations, translation, and stowage placement in the training facilities. Given the constraints of a 1g-based translation path, it is not possible or safe to place stowage items where they would potentially be stowed onboard the ISS. On-orbit, there is the benefit of weightlessness, which allows stowage of items on any axis with proper restraint. The crew can translate through the available volume and position their bodies to move around obstructions or protrusions in the translation paths. Additionally, while on-orbit, some of the stowage lockers are tightly packed, making it difficult to re-stow items, due to the lack of gravity working against the crewmember. Similarly, crewmembers often have trouble with items floating off during retrieval or re-stowage. In a 1g environment, stowage does not behave the same as in a 0g environment. However, tasks and procedures may be written based on what is known from testing in a 1g environment and training in a mock-up to a best representation of what it will be like for the crewmembers while on-orbit. Given the gravitational differences between Earth and orbit, as well as the operational disconnect between ground training and life on-orbit, the crewmembers often experience steep learning curves once onboard the ISS. The result is that upon arrival onboard the ISS, the crew often has difficulty managing stowage and operating nominally due to an initial lack of technical knowledge of on-orbit configurations paired with 0g conditions, and errors can result because the simulators and mockups do not provide adequate information about the 0g conditions.

A reason for failure to recall procedural knowledge that crewmembers were trained on is extreme circumstances, such as off-nominal conditions, stress, and high workload. The Russian spacecraft Progress 234 collided with the Mir space station, causing the pressure hull to rupture and nearly causing the Mir to be abandoned (Shayler, 2000; Ellis, 2000). The original operations concept was that Progress would dock automatically with the Mir. The primary responsibility of the crew was to cancel the approach, but very limited crew displays were provided. Crewmembers were later tasked to perform the docking because building automatic docking systems for each Progress spacecraft was undesirable. A number of contributing factors were determined for this off-nominal incident, including the vehicle condition and shutdown of the Kurs radar system during the Progress 234 docking, which deprived the crew of necessary range data. It was determined that the crash had three immediate causes: an initial closing rate that was higher than planned, late realization that the closing rate was too high, and incorrect final avoidance maneuvering. Several types of human factors task design issues may have contributed to this incident, such as psychophysical, sensory-motor, and cognitive factors (Ellis, 2000). The crew also experienced stress because of excessive workload and repeated system failures, which continuously commanded their attention and, in turn, reduced vigilance on their primary task (Ellis, 2000). In addition, and specifically relevant to technical and procedural knowledge, the last formal training the crew received was 4 months before the docking event, and they may not have had sufficient or timely practice in task design to handle the conditions. The aforementioned factors resulted from a poor operations concept and an absence of systematic analysis of the tasks to be performed and the information (inputs) and controls (outputs) that were necessary for the crew to perform them.

VI. Computer-Based Modeling and Simulation

Understanding and predicting human-system performance and identifying risks that may be inherent in a concept or a design is often achieved via computer-based modeling or simulation. The use of human performance models can result in significant lifecycle cost savings compared with repeated human-in-the-loop (HITL) evaluations, but accurately modeling the human is extremely difficult. Modeling and simulation can be used to understand specific tasks and workload levied on human operators; however, these kinds of modeling and simulation methodologies are not readily available. In the Space Human Factors Engineering (SHFE) domain, modeling and HITL evaluations must be used in concert. We do not have high-fidelity human performance models (HPMs), and most of the existing models have not been sufficiently validated or certified. Accordingly, models must be used in a limited fashion, i.e., to help determine the critical areas that should be addressed through the more costly, but more representative, HITL evaluations.

A. Human Performance Models and Workload

A model is a general term that encompasses many forms, ranging from static paper models of conceptual flow patterns to conceptual mathematical representations of the basic human processes (e.g., memory, attention) and to dynamic and integrated computationally generated predictive models of operator-environment interaction generating workload and timing output (among other measures). Human performance modeling is the process whereby human characteristics are embedded within a computer software structure to represent a “virtual” human operator interacting with a “virtual” operating environment. Integrated HPMs simulate and predict emergent behavior based on multiple, interacting sub-models of human behavior, such as perception, attention, working memory, long-term memory, and decision-making. This is accomplished typically by symbolically incorporating sub-models of human performance that feed both forward and back to other constituent models within the human information processing system. The use of appropriate and validated integrated HPMs can thus support the basic human factors principle of predicting the impact of alternative design options early in the system design process. These HPMs may also be used synergistically with HITL studies, especially when complex systems are being developed, a time when events cannot be studied fully with HITL subjects due to safety concerns, cost considerations, or practical difficulties associated with the simulation of very rare events (Gore and Corker, 1999). Modeling can provide evidence for workload being a contributing factor to inadequate task design. NASA and the Federal Aviation Administration (FAA) collaborated to develop and use human performance models to predict human-system interactions.

B. *Predictive Modeling of Operator Workload*

The challenge associated with the measurement and management of workload from an empirical perspective has led to many different conceptualizations on the degree to which workload should influence an operator's performance. There is little question that workload does impact nominal performance, but there is less agreement on precisely how workload influences performance. Some individuals thrive under periods of high task load and fail under very low workload, while others experience the opposite effect. Representing this divergent empirical performance computationally is needed so that accurate representations of human-system interactions are generated by model analysts. The Man-machine Integration Design and Analysis System (MIDAS), a NASA Ames software tool, possesses two distinct approaches to represent workload and its impact on operator performance, both of which rely on the Multiple Resource Theory (MRT) operating behind the scenes to impact performance. The first is an unconstrained representation where MIDAS completes the tasks and outputs workload according to a conflict matrix with task degradation functions based on the MIDAS behavioral primitives linked to the task. As there are no upper limits or redlines on operator performance, this mode of operation allows model analysts to see where the model predicts workload spikes. The second approach is a representation that includes a task management model that uses task priority to determine task schedule.

C. *MIDAS Representation of Workload*

The MIDAS workload model computes the workload of a multi-tasking operator along seven channels: two input channels, Visual and Auditory; two cognitive processing channels, Cognitive-Spatial and Cognitive-Verbal; and three output channels, Voice, Gross Motor, and Fine Motor. The MIDAS workload model is based on the Task Analysis/Workload (TAWL) task-load specification structure, which was developed and validated by the U.S. Army

(McCracken and Aldrich, 1984; Hamilton, Bierbaum, and Fulford, 1991; Hamilton, Bierbaum, and McAnulty, 1994; Mitchell, 2000). Basic behavioral primitives are created and workload estimates are assigned to the behavioral primitives using the TAWL. These behavioral primitives are then tied to MIDAS tasks in the task network model that are then exercised in various operational contexts. The MIDAS workload model also contains the MRT (Wickens, 1984; Wickens, 2002) of human attention for situations where multiple tasks share attentional resources. In his 2002 article, Wickens outlined that the MRT must possesses four features. The first is that each task can be represented as a vector of its resource demands at a qualitative (type of resource) and a quantitative (number of resources) level. Second, channel load is task-dependent. Third, loss of performance on one or both tasks from its single task level is determined by a performance penalty if: (a) the total demand on both tasks is high and (b) both tasks compete for overlapping resources (common levels on one of the dichotomous dimensions) within the four dimensions of the multiple resource model (or within the dimensions of whatever other model is selected). Fourth, the amount of performance lost on the tasks can be established by an allocation policy. If both tasks have equal priority, each task will share the performance decrement.

D. *Human Performance Model Prediction of Workload*

The MIDAS v5 (Gore et al., 2009; Gore, 2008; Gore et al., 2008) was adapted to model flight crew performance with augmented automation characteristic of the Next Generation Air Transportation System (NextGen) operations (Gore, 2010; Joint Planning and Development Office [JPDO], 2009). In a recent MIDAS v5 research study, a high-fidelity model of a two-pilot commercial crew flying approach and landing operations was developed and validated (Gore, Hooey, Haan, Bakowski, and Mahlsted, 2011; Gore, Hooey, Socash et al., 2011). The model inputs, including the task trace and input parameters, were validated using focus group sessions comprised of a total of eight commercial pilots with glass-cockpit aircraft and area navigation flying experience. The pilot-centric scenario-based cognitive walkthrough approach captured the context of operations from 10,000 feet to touchdown and enabled pilots to assess the modeled tasks and identify tasks that were missing or in the wrong sequence. Over 1100 tasks were encoded into this HPM. The pilots also completed quantitative rating scales, which were used to validate the model input parameters for workload and visual attention. The model was refined based on the results of this input validation process. Next, the model outputs, workload and visual attention, of the refined model were statistically compared to existing HITL data. The workload model output correlated with a comparable HITL study with an r^2 of 0.54 for overall workload. The individual workload dimensions also correlated positively with the HITL study, with the r^2 ranging from 0.55 to 0.94 and from 0.55 to 0.94 for the individual workload dimensions. Visual percent dwell time correlated with three HITL studies with $r^2 = 0.99$. These validation results provide confidence that the model validly represents pilot performance.

The MIDAS closely spaced parallel operations (CSPO) scenarios were then used to evaluate proposed changes to flight deck technologies, pilot procedures, operations, and roles and responsibilities to support the development of the Next Generation (NextGen) CSPO technologies and concepts and to explore “what-if” scenarios about the impact of the current day, a transition timeframe, and NextGen CSPO concept on pilot performance in response to off-nominal events. Compared to conditions when the controllers were responsible for separation in the descent and landing phases of flight, the output from these model predictions suggests that the response time of the flight deck to detect the lead aircraft blunder will decrease, pilot scans to the navigation display will increase, and workload will increase. A number of “what-if” scenarios were also assessed based on the information required depending on the roles and responsibilities under evaluation. All of the analyses provided NASA with documented, valid procedural sets that have been encoded into a computational framework (see Gore, Hooey, and Foyle, 2012). It is anticipated that these procedural models will be re-used in future phases of the NextGen research. Model results such as these are critical for estimating human performance with automation (as illustrated by the automation required for NextGen operations) and for estimating workload as the operator interacts with NextGen displays and technologies.

E. *Augment and Refine Workload Primitives within Computational Models for Long-Duration Space Operations*

The iterative approach to develop valid workload primitives illustrated by the MIDAS FAA simulation is an example of the approach that needs to be adhered to when developing credible human performance models for the microgravity and space environment, particularly those that use workload as an output. There is a need for further work on the workload primitives to ensure that they correctly represent the workload of operators in space.

VII. Risk in the Context of Exploration Mission Operational Scenarios

Future exploration mission scenarios will increase in duration and in distance from Earth. This will require developing new technology, new work methods, and new ways of ensuring that these novel elements are suitably integrated. Missions carried out in space will need greater flexibility and less dependence on ground support, and new interaction between ground-based resources and crew will also be needed. Designing spacecrafts to be operated in low-Earth orbit or on the lunar surface, where communications may be interrupted by malfunctions, requires careful design, particularly for the tasks needed to maintain the spacecraft and complete the mission objectives. Knowledge can be provided from other sources, such as onboard databases or mission support on Earth. However, tasks cannot require more immediate knowledge than can be accessed by four individuals. This very low number of humans available for any task will require the careful use of automation to augment their capabilities. Routine tasks that consume a great deal of time can be automated to free crewmembers for performing the tasks that require uniquely human judgment, creativity, and problem solving skills.

A specific requirement for increased autonomy for lunar and Mars missions is automated planning capabilities and tools. These tools would provide necessary automated support to determine alternative plans and solutions for managing daily tasks. Current spaceflight crews rely on onboard, automated systems, and future crews with increased flight duration and increased autonomy will rely even more on these systems to provide information that is appropriate, accurate, and up to date. Increased automation will result in the need for additional training to ensure that the crew can perform the automated tasks in the event of automation failure. Automated tasks must be carefully designed to prevent the crew from becoming unaware of or complacent about potential hazards. This situation could ultimately result in system errors, degraded crew performance, and compromised crew and vehicle safety.

VIII. Gaps

Measuring workload is a challenge because of the need to match the workload measurement method with the evaluation objectives. As a result, measuring workload for a "system", a set of activities, groups of people, and/or an entire mission remains difficult. Performance data related to workload and data describing the relationship between efficiency, task design, and workload in the context of spaceflight and long-duration missions are not available. Research needs to be done in the area of unobtrusive methods and tools to measure workload in situ in the context of long-duration spaceflight. Technologies to measure workload objectively should be studied in particular. Currently, the consensus in the workload research community is that available objective measures of workload require too much equipment and are obtrusive. Workload measures can interrupt and change the ongoing tasks, thus changing the workload level of the human. The relationship between performance and workload is not clear, and more studies and analysis are needed to determine under what conditions high or low workload has a negative effect on performance. Schedule and length of work day also affect workload, and there is little to no research data on these factors.

Identifying and anticipating potential performance breakdowns that result from workload imbalances due to improper task allocation among the human, robot, and human-robot team. A failure of any part of this human-robot system will have widespread consequences for the efficiency and safety of the overall system. By surveying the knowledge base in human-automation interaction, the ability of designers to predict and avoid breakdowns in human and overall system performance will be improved.

The challenge in predicting workload's impact on operator performance stems from the difficulty in correctly identifying precursors to task failure; the impact that pending tasks have on operator performance; the impact of workload on vigilance levels/boredom, queue length, task load, and task complexity; and successful task scheduling and task performance. Some of the above variables are driven by the system (e.g., task load and queue length), while others emerge from the Mission Specialist's performance (e.g., task scheduling ability). The challenge outlined above is increased when applying workload's impact at the system level because of the increased number of free parameters within a system that are not considered at the individual level. This challenge is increased further when the human's interaction with automation (and by extension robots) is considered.

Research has shown that automation can reduce operator workload and fatigue. Furthermore, when automation is introduced, the nature of workload shifts from motor (e.g., manual control) to cognitive (e.g., supervisory tasks such as monitoring, projecting, and analyzing) (Parasuraman and Hancock, 2007; Sheridan and Parasuraman, 2000; Billings, 1997). When patterns of workload shift in such ways, new system vulnerabilities arise (NRC, 1993). As the automation becomes more pervasive, such vulnerabilities may contribute to safety risks.

Casner and Gore (2010) identified that the workload measures and techniques have been designed primarily for aviation contexts and have been applied to other highly procedural domains examining tasks over short periods of time (Hart, 1978; Hart and Staveland, 1988; Casner, 2005; Hancock and Chignell, 1988; Stein, 1985; Reid, 1989; Gawron, 2008). The measures have been designed and used primarily at the individual level and not extended to team operations and certainly less so to a human-robot team (note, research does exist on human-robot interaction but this is applied at the level of robot display design and its effect on human operation or on the impact of teleoperation on human performance; see Prewett et al., 2010). Operations in space, however, will involve extensive human-robot team performance and will require extended missions lasting upwards of 30 days with multiple human and robot crew members sharing in the task performance, often simultaneously.

Potential gaps related to inadequate task design include, but are not limited to:

- Inadequate methodologies and techniques for developing procedures for long-duration missions when ground support will not be as readily available as for current space missions.
- Lack of research on the extent to which procedures should be automated for long-duration missions and how this automation may change the crewmembers' task.
- Lack of knowledge regarding procedure development and just-in-time training for long-duration missions.

- Little to no documented data on crewmember retention of procedural task steps presented pre-flight and then performed during a mission.
- Lack of human modeling methods that could be used to study and predict workload and procedural guidance-related phenomena.
- Task designs that optimize workload for human-human tasks, human-robot tasks, and robotic tasks.
- The issue of task shedding as a way to deal with periods of high workload (e.g., consider the multiple resource model and the notion of workload redlines; Wickens, 2010).

Potential gaps related to human performance models:

There is no HPM that accounts for the skill degradation that might occur as a result of the time between the training and the required human performance or that properly characterizes the astronauts' autonomy (and possible lack of knowledge) experienced during longer duration missions. Most HPMs developed to date have been in the area of military, commercial aviation, and, to a lesser extent, surface transportation (Gore, 2010). Most procedures developed to date assume that there will be a life-line to ground control for emergency procedures and problem solving or that there will be specialized crew onboard to resolve an anomaly. Such an assumption will likely not hold for longer duration missions due to planet shadows or bandwidth limitations that occur given the distance between ground control and the vehicle. Since it will be unlikely to count on just-in-time training and since longitudinal empirical studies are rare and very costly to conduct (e.g., Mars analog study, Dinges et al., 2011, Desert Research and Technology Studies, NASA Extreme Environment Mission Operations), HPMs are needed to estimate the impact that longer duration missions will have on the human operator, particularly in the response to off-nominal and unexpected conditions.

It is critical to ensure that training for critical task response is current despite months between initial training and skill utilization and in the event that the just-in-time training scripts that are uploaded to the crew are interrupted (by a bandwidth limit). As such, HPMs are needed that accurately reflect:

- Skill degradation for likely critical tasks conducted in long-duration missions
- The effect of inadequately maintaining skills: (a) candidate tools are needed to test alternate automation design approaches and (b) candidate what-if scenarios are needed to train crews in how to more effectively respond to flight-critical tasks
- Long-duration mission performance variables likely to impact crew performance (e.g., social issues, boredom, fatigue, and vigilance; see Hancock in Gore, Macramalla, and Ahumada, 2010)

- The impact on the crew's maintenance of skills given more variable time delays (variable windows of opportunity) characteristic of longer duration missions

Potential gaps related to workload measurement:

Current approaches (e.g., subjective HITL measures) to study long-duration workload that results from critical task design remain rooted in the established measurement methods and are applicable mainly to short-duration operations. Extrapolating the short-duration-determined workload measurement methods to long-duration mission operations may not provide accurate insights into workload and its impact on long-duration mission operation success. As such, gaps in critical task design knowledge that should be considered for this risk include, and research and literature reviews will be solicited in these areas:

- Validation of workload metrics as predictors of long-term mission task accomplishment (including identifying and categorizing behavioral markers)
- Possible trade-offs between paper and electronic media (what are the criteria for media use in task aids to support task procedures) for critical task designs
- Workload metrics for multiple operators in long-duration mission operations, as situations with multiple operators will often require collaborative problem solving
- Effects of longer and more variable time delays on task execution on long-duration missions
- Lack of computational models of long-duration mission operations, task designs, and workload
- Relationship between robotic operations, automation, and workload
- Augment and refine workload primitives that are used for critical tasks within computational models for long-duration space operations

A summary of all SHFE gaps can be found in the Human Research Roadmap Content Management System at http://sa.jsc.nasa.gov/hrrcms/.

IX. Conclusion

The Risk of Inadequate Critical Task Design relates to the definition and development of mission tasks and to the interactions between multiple crewmembers, crew and robotics/automation, and crew and ground-control personnel. Workload is an important contributing factor in crew interaction with systems. Designers must consider workload when designing hardware and software with crew interfaces, procedures, and operations. Procedural guidance is a factor when written direction, checklists, graphic depictions, tables, charts, or other published guidance is inadequate, misleading, or inappropriate, creating an unsafe situation. Technical and procedural knowledge is a factor when an individual was adequately exposed to the information needed to perform the mission element but did not absorb it. Lack of knowledge implies no deficiency in the training program, but rather the failure of the individual to absorb or retain the information. Exposure to information at a point in the past does not imply "knowledge" of it.

Substantial evidence supports the claim that inadequate task design leads to performance errors, which in turn leads to decreased safety and efficiency. Task analysis; operations concepts; appropriate function allocation between humans, spacecraft systems, and habitats; and assessment of integrated human-system performance are key components in avoiding this risk. Loukopoulos, Dismukes, and Barshi (2009) demonstrated that even a task as carefully developed and refined over many years as operating an aircraft can be significantly improved by a systematic analysis. Unfortunately, a systematic analysis of task procedures rarely occurs during the preliminary design phase, when modifications are most feasible. Although operational procedures for task performance by spaceflight crews are executed in mockups and simulators during preflight training and feedback regarding tasks and operations is received, hardware, system, and software design is often relatively mature. Therefore, it is often cost prohibitive to modify design based on feedback from training related to tasks and operations. As a result, task performance is often negatively impacted in terms of efficiency and crew time.

In order to close the above-mentioned gaps, there is a need for operational onboard data collection on workload, usability, task design, and performance related to scheduling for model and simulation development and validation for human spaceflight operations in order to capture information that will aid in preparations for long-duration missions.

X. References

AAIB. (1977). Aircraft Accident Report 5/77 - British Airways Trident G-AWZT Inex Adria DC9 YU-AJR, Report on the collision in the Zagreb area, Yugoslavia, on 10 September 1976. London: Federal Civil Aviation Administration, Aircraft Accident Investigation Commission

Abbott, K., (2010). Presentation made at the Flight Safety Foundation International Aviation Safety Seminar, November, 2010, Milan, Italy.

Ainsworth, L.K. (2004). *Task analysis*. In Sandom, C., Harvey, R. (Eds.), Human Factors for Engineers (81–112). London: Institution of Engineering and Technology.

ALPA (1977). Human factors report on the Tenerife accident. Engineering and air safety. Washington DC.

Baddeley, A (1992). Working memory. *Science* 255(5044): 556–9.

Baggerman, S.D., Rando, C., Duvall, L.E. (2004) Habitability and human factors: lessons learned in long duration spaceflight. In: *Proceedings of the American Institute of Aeronautics and Astronautics Space 2004 Conference Exhibit*. San Diego, CA, Sep. 2004.

Billings, C. (1997). Aviation automation: The search for a human-centered approach. Mahwah, NJ: Erlbaum.

Casner, S. M., & Gore, B. F. (2010). *Measuring and evaluating workload: A primer*. (NASA/TM 2010-1850). Washington, D.C.: National Aeronautics and Space Administration.

Casner, S.M. (2005) Transfer of learning between a small technically advanced aircraft and a modern commercial jet. *International Journal of Applied Aviation Studies 5 (2), 307-319*

Desaulniers, D. R., Gillan, D. J., & Rudisill, M. (1988). *The effects of format in computer-based procedure displays.* Proceedings of the Human Factors and Ergonomics Society Meeting.

Dinges, D.F., Mollicone, D.J., Ecker, A.E., Metaxas,D., Basner, M. (2011). Objective monitoring of crew neurobehavioral functions in IBMP's 520-day Mars simulation. Poster session presented at the 18th IAA Humans in Space Symposium, Houston TX, 2011.

Ellis, S. R. (2000). Collision in space. Ergonomics in Design: The Quarterly of Human Factors Applications, 8(1), 4-9.

Gacy, A.M., Wickens C.D., Sebok, A., Gore, B.F., Hooey, B.L. (2011). Modeling Operator Performance and Cognition in Robotic Missions. In the *Proceedings of the 55th Annual Meeting of the Human Factors and Ergonomics Society*, Las Vegas, NV, September 19-23.

Gawron, V.J. (2000). Human performance measures handbook: CRC Press.

Gawron, V.J. (2008). *Human performance, workload, and situational awareness measures handbook* (2nd ed.). Boca Raton, Florida: CRC Press, Taylor & Francis Group.

Gore, B.F. (2002). Human performance cognitive-behavioral modeling: A benefit for occupational safety. *International Journal of Occupational Safety and Ergonomics* (JOSE), 8 (3), 339-351.

Gore, B.F. (2008). Human performance: Evaluating the cognitive aspects. In V. Duffy (Ed.), *Handbook of Digital Human Modeling* (Chapter 32). NJ: Taylor and Francis.

Gore, B.F. (2010). Man-machine integration design and analysis system (MIDAS) v5: Augmentations, motivations, and directions for aeronautics applications. In P.C. Cacciabu, M. Hjalmdahl, A. Luedtke & C. Riccioli (Eds.), *Human Modeling in Assisted Transportation*. Heidelberg: Springer.

Gore, B.F. Macramalla, S., Ahumada, A., & Oyung, R. (2010). Workload Considerations in Long Duration Space Operations: A System Perspective. Final Milestone Report Directed Research Project: Workload. Moffett Field CA: NASA Ames Research Center.

Gore, B.F., & Corker, K. (1999, 1999-05-18). *System interaction in free flight: A modeling tool cross comparison.* Paper presented at the International Conference on Digital Human Modeling for Engineering and Design, The Hague, NL.

Gore, B.F., Hooey, B.L., & Foyle, D.C. (2012). A Validated Set of MIDAS v5 Task Network Model Scenarios to Evaluate NextGen Closely Spaced Parallel Approach Concepts, In Human Centered Systems Lab (Ed.), *HCSL Technical Report (HCSL-12-03)*. Moffett Field, CA: NASA Ames Research Center.

Gore, B.F., Hooey, B.L., Foyle, D.C., & Scott-Nash, S. (2008, July 14-17). *Meeting the challenge of cognitive human performance model interpretability though transparency: MIDAS v5.x.* Paper presented at the Applied Human Factors and Ergonomics International Conference, Las Vegas, Nevada.

Gore, B.F., Hooey, B.L., Haan, N., Bakowski, D.L., & Mahlsted, E. (2011, July 9 - July 14). *A methodical approach for developing valid human performance models of flight deck operations.* Paper presented at the 14th Annual HCI International 2011, Orlando, FL.

Gore, B.F., Hooey, B.L., Mahlstedt, E., & Foyle, D.C. (2012). Evaluating NextGen Closely Spaced Parallel Approach Concepts with Validated Human Performance Models: Model Development, Implementation, and Results. In Human Centered Systems Lab (Ed.), *HCSL Technical Report (HCSL-12-04)*. Moffett Field, CA: NASA Ames Research Center.

Gore, B.F., Hooey, B.L., Mahlstedt, E., & Foyle, D.C. (2012). Extending validated human performance models to explore NextGen Concepts. In the *4th International Conference on Applied Human Factors and Ergonomics (AHFE),* July 2012.

Gore, B.F., Hooey, B.L., Socash, C., Haan, N., Mahlsted, E., Bakowski, D.L., Foyle, D.C. (2011). Evaluating NextGen closely spaced parallel operations concepts with human performance models (HCSL, Trans.) *HCSL Technical Report (HCSL-11-01)*. Moffett Field, CA: NASA Ames Research Center.

Gore, B.F., Hooey, B.L., Wickens, C.D., & Scott-Nash, S. (2009). A computational implementation of a human attention guiding mechanism in MIDAS v5. In V. G. Duffy (Ed.), *Digital Human Modeling* (Vol. 5620/2009, pp. 237-246). Berlin / Heidelberg: Springer.

Guhe, M., Liao, W., Zhu, Z., Ji, Q., Gray, W. D., & Schoelles, M. J. (2005). *Non-intrusive measurement of workload in real-time*. Paper presented at the 49th Annual Meeting of the Human Factors and Ergonomics Society.

Hamilton, D. B., Bierbaum, C. R., & Fullford, L. A. (1990). Task analysis/workload (TAWL) user's guide, version 4 (Research Project 91-11). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences (AD A241 861).

Hamilton, D.B, Bierbaum, C.R., & McAnulty, D.M. (1994). Operator task analysis and workload prediction model of the AH-64D mission volume II: Appendices A through F (Contract # MDA90-92-D-0025). Army research Laboratory. Anacapa Sciences, Inc., Fort Rucker, AL, USA.

Hancock, P.A., & Chignell, M.H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man, and Cybernetics, Vol. 18,* No. 4, 647-659, July/August

Hancock, P.A., & Szalma, J.L. (2008). *Chapter 1: Stress and performance*. In P. A. Hancock & J.L. Szalma (eds.), Performance Under Stress (pp.1-18), Ashgate Publishing Burlington, Vermont.

Hart, S. G., & Staveland, L. (1988). Development of the NASA task load index (TLX): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 239-250). Amsterdam: North Holland.

Hart, S., and Wickens, C.D. (2008) Mental Workload. In NASA Human Integration Design Handbook; National Aeronautics and Space Administration, Washington, DC.

Heymann, M., Degani, A., & Barshi, I. (2007). Proceedings of the 14th International Symposium of Aviation Psychology: *Generating Procedures and Recovery Sequences: A Formal Approach*, Dayton, OH.

Hill, S. G., Iaveccia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L., & Christ, R. E. (1992). Comparison of four subjective workload rating scales. *Human Factors, 34*, 429-439.

Hooey, B. L., & Foyle, D. C. (2008). Advancing the state of the art of human performance models to improve aviation safety. In B. L. Hooey & D. C. Foyle (Eds.), *Human performance modeling in aviation* (pp. 321-349). Boca Raton: CRC Press/Taylor & Francis.

Hudy, C., & Woolford, B. (2008). Space human factors engineering gap analysis project final report. National Aeronautics and Space Administration; 2007 [cited 2008 Feb 12]. Available from: http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20070020456_2007018857.pdf.

Hudy, C., Byrne, V., Smith, D., and Whitmore, M. (2005). Human Factors Assessment of Respiratory Support Pack (RSP) Cue Card. Proceedings of the Aerospace Medical Association Annual Meeting, Kansas City, MO.

ISO (1999). International Standard for Human-Centred Design Processes for Interactive Systems. ISO 13407

JPDO, (2009) Concept of operations for the next generation air transportation system. *Vol. 3*. Washington, DC: JPDO.

Kirwan, B., & Ainsworth, L. K. (Eds.). (1992). *A guide to task analysis*. Philadelphia, PA: Taylor & Francis.

Kranz, G. (2000). *Failure is not an Option*. New York, New York: Simon & Schuster.

Ladkin, P. (1996). AA965 Cali Accident Report. *University of Bielefeld.*

Lee, P. (2002). Mars on Earth. *Ad Astra*, *14* (3), 12–17, 51–53.

Legner, K. (2004). Humans in Space & Space Biology. Seminars of the United Nations Programme on Space Applications. Selected Papers from Activities Held in 2003. United Nations, New York, 2004. ISBN 92-1-100935-9, p.77-135

Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2009). *The Multitasking Myth: Handling Complexity in Real-World Operations*. Surry, England: Ashgate Publishing Limited.

McCracken, J. H., & Aldrich, T. B. (1984). Analysis of selected LHX mission functions: Implications for operator workload and system automation goals (Technical note ASI 479-024-84[b]). Fort Rucker, AL: Anacapa Sciences, Inc.

McCurdy, M., Connors, C., Pyrzak, G., Kanefsky, R., & Vera, A. H. (Eds.). (2006) Proceedings of the Conference on Human Factors in Computing Systems CHI'06: *Breaking Through the Fidelity Barriers: An Examination of Our Current Characterization of Prototypes and an Example of a Mixed-Fidelity Success*. Montreal, CA.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63 (2); 81-97.

Mitchell, D.K. (2000). *Mental workload and ARL workload modeling tools*. ARL-TN-161. U.S. Army Research Laboratory: Aberdeen Proving Ground, MD.

NRC (1993). *Workload Transition: Implications for Individual and Team Performance*. In B.M. Huey & C.D. Wickens, (eds) National Research Council Panel on Workload Transitions. Washington DC: NAP.

NTSB. (1987). Aircraft Accident Report- Northwest Airlines, Inc. McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne Country Airport, Romulus, Michigan, August 16, 1987 (No. NTSB-AAR-88/05), Washington DC.

Parasuraman, R., & Hancock, P.A. (2007). *Mitigating the adverse effects of workload, stress, and fatigue with adaptive automation.* In P.A. Hancock & J.L. Szalma (eds), Performance Under Stress (pp. 45-57). Burlington, VT: Ashgate.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced complacency. *International Journal of Aviation Psychology, 3*(1), 1-23.

Prewett, M.S., Johnson, R.C., Saboe, K.N, Elliott, L.R., Coovert, M.D. (2010). Managing workload in human–robot interaction: A review of empirical studies. Computers in Human Behavior, 26(5), September, 2010, 840-856

Rando C., Baggerman S.D., Duvall L.E. (2005) Habitability in space. In: *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting,* Orlando, Fla., Sep 26–30, 2005, pp. 5–9.

Roscoe, A. H., & Ellis, G. A. (1990). A subjective rating scale for assessing pilot workload in flight: A decade of practical use. Farnborough, UK: Royal Aerospace Establishment.

Salas, E., Wilson, K., Priest, H., & Guthrie, J. (2006). *Design, delivery, and evaluation of training systems.* In G. Salvendy (Ed.), Handbook of human factors and ergonomics, John Wiley & Sons; (472–512).

Sarter, N. B. & Woods, D. D. (1997). Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. Human Factors, 39, 553-569;

Sarter, N. B., Woods, D.D. (1995). How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors,* March 1995, 37: 5-19.

Sebok, A., Wickens, C. D., Gacy, A. M., Brehon, M., Scott-Nash, S., Sarter, N., Li, H. Gore, B. F., & Hooey, B. L. (2012, July 10-14, 2012). *MIDAS-FAST: A Modeling and Simulation Based Tool to Predict Operator Performance in Human-Robotic Automation Systems.* Presented at the Applied Human Factors and Ergonomics 2012, San Francisco, CA.

Shappell, S.A., and Wiegmann, D.A. (2000). “The Human Factors Analysis and Classification System – HFACS”, Report DOT/FAA/AM-00/7, February 2000. (http://www.nifc.gov/fireInfo/fireInfo_documents/humanfactors_classAnly.pdf).

Shayler, D.J. (2000). *Disasters and Accidents in Manned Spaceflight*. Chichester, UK: Springer-Praxis Publishing.

Sheridan, T. (Ed.). (2002). Human and automation: Systems design and research issues. New York: Wiley.

Sheridan, T. B., & Parasuraman, R. (2000). Human versus automation in responding to failures: an expected-value analysis. [Comparative Study]. *Human Factors, 42*(3), 403-407.

Stein, E.S. (1985). *Air traffic controller workload: An examination of workload probe*. (Report No. DOT/FAA/CT-TN84/24). Atlantic City, NJ: Federal Aviation Administration Technical Center.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). *Workload and vigilance*. Paper presented at the Human Factors and Ergonomics Society 42nd Annual Meeting, Chicago, Illinois.

Wickens, C. D. (1984). The multiple resource model of human performance: Implications for display design. Williamsburg, VA: AGARD/NATO Proceedings.

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3(2), 159–177.

Wickens, C. D. (2010). Mental workload prediction and the redline. In B. Gore, S. Macramalla, & E. Salud (eds.), *A workshop on workload scales, measurement, and management for long duration space operations* (No. NASA/CP-2010-216398). Washington, D.C.: National Aeronautics and Space Administration.

Wickens, C. D., Li, H., Santamaria, A., Sebok, A., & Sarter, N. B. (2010). *Stages and levels of automation: An Integrated meta-analysis.* Paper presented at the Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco.

Wickens, C.D. (2008). Multiple resources and mental workload. Human Factors, 3 (2), 159–177.

Wiegmann, D., Faaborg, T., Boquet, A., Detwiler, C., Halcomb, K., & Shappell, C. (2005). *Human error and general aviation accidents: A comprehensive, fine-grained analysis using HFACS* (Final Report No. Technical Report DOT/FAA/AM-05/24). Office of Aerospace Medicine, Federal Aviation Administration: Washington, DC.

XI. Team

Kritina L. Holden, Ph.D.	Lockheed Martin, NASA Johnson Space Center
Vicky Byrne, M.S.	Lockheed Martin, NASA Johnson Space Center

XII. List of Acronyms

AAIB	Air Accidents Investigation Branch
ALPA	Air Line Pilots Association
CSPO	Closely Spaced Parallel Operations
DoD	Department of Defense
FAA	Federal Aviation Administration
FAST	Future Attribute Screening Technology
FCI	Flight Crew Integration
HCD	Human-Centered Design
HCI	Risk of Inadequate Human-Computer Interaction
HFACS	Human Factors and Analysis Classification System
HITL	Human-in-the-loop
HMD	Head Mounted Display
HPM	Human Performance Model
HRP	Human Research Program
IMG	Integrated Medical Group
ISO	International Standards Organization
ISS	International Space Station
JIT	Just-In-Time
JPDO	Joint Planning and Development Office
KLM	Koninklijke Luchtvaart Maatschappij (Royal Dutch Airlines)
MIDAS	Man-Machine Integration Design and Analysis System
MRT	Multiple Resource Theory
NASA	National Aeronautics and Space Administration
NextGen	Next Generation
NRC	National Research Council
NTSB	National Transportation Safety Board
OpsCon	Operations Concept
PDA	Personal Digital Assistant
RMAT	Risk Management and Analysis Tool
RSP	Respiratory Support Pack
SA	Situation Awareness
SHFE	Space Human Factors Engineering
SLEEP	Risk of Performance Errors Due to Fatigue Resulting from Sleep Loss, Circadian Desynchronization, Extended Wakefulness, and Work Overload
TASK	Risk of Inadequate Critical Task Design
TAWL	Task Analysis/Workload
TRAIN	Risk of Performance Errors Due to Training Deficiencies