AN ABSTRACT OF THE THESIS OF

Jennifer Rae Wilson for the degree of Master of Science in Industrial Engineering presented on February 13, 1998. Title: The Effect of Automation on the Frequency of Task Prioritization Errors on Commercial Aircraft Flight Decks: An ASRS Incident Report Study.

Redacted for Privacy

Abstract approved: ______________________________
Kenneth H. Funk, II

Task Management (TM) refers to the function in which the human operator manages his/her available sensory and mental resources in a dynamic, complex, safety-critical environment in order to accomplish the multiple tasks competing for a limited quantity of attention. There is reason to believe that the level of automation on the commercial aircraft flight deck may effect TM, however to date there has been little research that directly addresses this effect. Thus, the primary objective of this study was to begin evaluating the relationship between TM of commercial airline pilots and the level of automation on the flight deck by determining how automation affects the frequency of Task Prioritization errors as reported in Aviation Safety Reporting System (ASRS) incident reports. The secondary objective of this study was to create a methodology that modeled an effective way to use ASRS incident report data in an inferential analysis.

Two samples of ASRS incident reports were compared. The first sample was composed of 210 incident reports submitted by pilots flying advanced technology aircraft and the second sample was composed of 210 incident reports submitted by pilots flying traditional technology aircraft. To help avoid confounding effects, the two samples were further divided into three sub-samples each made up of 70 reports submitted during a specified time period: 1988-1989, 1990-1991, and 1992-1993. Each incident report was analyzed using an incident analysis form designed specifically for this study. This form allowed the analyst to classify the incident report as either containing a Task Prioritization error or not based on the narrative of the report.

Twenty-eight incident reports from the advanced technology sample and 15 from the traditional technology sample were classified as containing Task Prioritization errors. Using the Chi Square (χ^2) test and a significance level of 0.05, this difference was found to be statistically significant.

The Effect of Automation on the Frequency of Task Prioritization Errors on Commercial Aircraft Flight Decks: An ASRS Incident Report Study

by

Jennifer Rae Wilson

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented February 13, 1998
Commencement June 1998

Master of Science thesis of Jennifer Rae Wilson presented on February 13, 1998

APPROVED:

Redacted for Privacy

Major Professor, representing Industrial Engineering

Redacted for Privacy

Head of Department of Industrial and Manufacturing Engineering

Redacted for Privacy

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Jennifer Rae Wilson, Author

AKNOWLEDGEMENTS

This work was supported in part by the Federal Aviation Administration, Office of the Chief Scientific and Technical Advisor for Human Factors (AAR-100) under grant 93-G-039 titled Comparative Analysis of Flight Deck Automation.

I am indebted to Dr. Ken Funk for giving me the opportunity to work with him in field of aviation human factors. The experience has truly changed my life. His guidance, support, and friendship have meant so much to me.

I would like to thank Dr. Ed McDowell for his invaluable advice regarding the statistical methods used in this study. In the end, I think his enthusiasm over performing statistical operations on the data began to rub off on me.

I would also like to extend my gratitude to Research Integrations, Inc. who hired me before this work was complete and then generously gave me the time and resources to finish writing this thesis.

A special thanks to Dr. Beth Lyall for her valuable contributions to this study, to Mary Niemczyk who helped me pull everything together in the end, and to Candy Suroteguh for his continued assistance throughout this project.

And, of course, I want to thank my Mom and my sister Pam. Your love, support, and belief in me have made it all possible.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

Figure

LIST OF TABLES

Table

The Effect of Automation on the Frequency of Task Prioritization Errors on Commercial Aircraft Flight Decks: An ASRS Incident Report Study

1. INTRODUCTION

1.1 Problem Statement

In recent years there has been a growing awareness of human factors issues associated with the increased presence of automated systems on modern commercial aircraft flight decks. As flight deck automation becomes more sophisticated, it is able to perform many of the tasks previously performed by the pilots and the pilot's role becomes more like that of a manager. With this shift in the pilot's role, new strategies must be developed for the pilots to successfully perform their tasks. In order to do this, the effect that automation has on pilot performance must be understood.

Technology has improved aircraft performance and reliability and by doing so has made significant contributions to both safety and the efficiency of operations. Despite this, there are still concerns about replacing the human functioning with automated systems because "it appears that the modern systems may be at the same time eliminating and producing errors; that certain types of errors are reduced, and others are enabled" (Wiener, 1989, p. 97). Enabling errors include both creating the possibility for new errors that did not exist previously and increasing the frequency of errors that already existed. Work focusing on these errors should not be to determine whether the traditional technology or the advanced technology produces the most errors overall, but rather to understand the errors that will be encountered in the new aircraft and how they may be successfully controlled.

While there is no comprehensive listing of the errors enabled by the advanced technology, there are a number of ideas about what they may be. For example, there is speculation that errors in Task Management may be enabled by the advanced technology.

Task Management (TM) refers to the function in which the human operator manages his/her available sensory and mental resources in a dynamic, complex, safety-critical environment in order to accomplish the multiple tasks competing for a limited quantity of attention. Flightcrews[1] must perform TM on the commercial flight deck because they do not possess the necessary resources to simultaneously execute all the tasks that demand their attention. The flightcrew must therefore prioritize the tasks in the order of most to least important and then allocate their resources according to this prioritization. In a dynamic system, the state of each task demanding attention continuously changes and as this occurs so too may change the relative urgency with which each task must be completed. Thus, the flightcrew must continuously perform the function of TM in order to maintain awareness of the changes in the state of the system and make the necessary revisions to the Task Prioritization. Flightcrews generally perform TM quite well as evidenced by the number of successful flights that occur each day. However, sometimes TM is not performed as effectively and the results can range from minor operational deviations to fatal accidents. Errors in TM can often be traced back to misprioritization. When a human operator misprioritizes tasks, I call this a Task Prioritization error.

There are several reasons why it has been speculated that TM errors may be enabled by the advanced technology. First, there are a greater number of tasks to be performed in the automated aircraft. All the flight control tasks found in the traditional technology aircraft must still be performed in the advanced technology aircraft but, in addition to these tasks, there are now tasks associated with communicating with and managing the automation. Adding tasks to the queue of tasks demanding attention increases the demands on the flightcrew. While the automation provides additional external resources for the flightcrew to utilize, these resources must be managed which increases demands on the function of TM. Second, the same resources may be overloaded in the automated aircraft. Some of the demands added by automation require

[1] In this study I have not differentiated between the roles of the individual pilots but instead have considered the human function at the level of the flightcrew. In this sense, the flightcrew in a 2-person flight deck has twice as many (human) resources available and twice as many resources to manage than does an individual pilot. Thus, I will refer to the human component as 'flightcrew' throughout this study. See Palmer, Hutchins, Ritter, & vanCleemput (1993) for a similar approach.

the cognitive processing resources that are already taxed in the traditional technology aircraft. Because of this, more prioritization may be required because more tasks are demanding the same resources. And third, some of the advanced systems, such as the Flight Management System (FMS), may inappropriately draw the attention of the flightcrew away from more critical tasks.

Because my interest lies in TM as it occurs in real flight operations, I ideally would like to collect the data from real flight operations. However, this method is often impractical. A viable alternative to viewing actual line operations is the use of incident reports submitted by pilots, such as those submitted to the Aviation Safety Reporting System (ASRS). ASRS incident reports are submitted voluntarily by aviation operations personnel (e.g., pilots, Air Traffic Controllers, flight attendants, ground personnel) and contain a description of a situation occurring in flight operations that the reporter believes has safety implications. With each report providing a description of an event that occurred in operations, they can be used as a practical way to view actual line operations from a pilot's perspective.

In the past due to the nature of the data, ASRS incident reports have been used primarily for descriptive analyses. In this study, however, it would be more useful to conduct an inferential analysis. Such an analysis may be conducted by carefully constructing a research question and choosing an appropriate statistical test. Because few researchers have taken this approach, there are not many examples of effective inferential analysis using ASRS incident report data.

1.2 Research Objectives

The flightcrew's function of TM on the commercial flight deck is an important part of flight operations, and committing errors in TM can have severe consequences. There is reason to believe that the level of automation may affect TM, however to date there has been little research that directly addresses this effect. Thus, the primary objective of this study was to begin evaluating the relationship between TM of commercial airline pilots and the level of automation on the flight deck by determining how automation affects the frequency of Task Prioritization errors as reported in Aviation Safety Reporting System (ASRS) incident reports.

Because ASRS incident reports are primarily used for descriptive analyses, a methodology for conducting a good statistical comparison analysis is lacking. Therefore, the secondary objective of this study was to create a methodology that models an effective way to use ASRS incident report data in an inferential analysis.

1.3 Overview of Thesis

The following is an overview of the remaining portions of this thesis.

Chapter 2 reviews the literature important to understanding how the level of automation present on the commercial flight deck could affect TM. First, flight deck automation and some of the issues that have been raised about it will be discussed. From this, it is evident that some basic concepts of human cognitive processing capabilities and limitations are important to understanding the flightcrew's performance on the flight deck, so the relevant research done in the area of attention addressing this will be presented. It is also important to understand how the flightcrew assesses a situation. Therefore, the concept of schema from cognitive science studies is presented to provide a valuable insight in how this may be done in a complex environment such as the commercial flight deck. Given these basic cognitive processing concepts, I will review some of the aviation human factors research concerned specifically with how a flightcrew manages tasks on the flight deck. Lastly, I will draw upon all these concepts to address the question: why would the level of automation on the flight deck affect task management?

Chapter 3 then presents a review of the literature responsible for the design of this study. This chapter first presents a description of the Aviation Safety Reporting System (ASRS) and its incident reports. Next, several valuable descriptive studies that have been performed using ASRS incident reports will be reviewed to illustrate the type of work and the findings that ASRS incident studies have yielded in the past.

Chapter 4 describes the methodology used in this study. This chapter describes the precautions taken to ensure that the study used a representative data sample and made a fair comparison between the advanced and traditional technology populations. It also includes a description of the analysis tool used to analyze the incident reports.

Chapter 5 presents the results of this study and includes the calculations that have been performed. Chapter 6 provides a discussion of the results and their relation to the objectives of this study and some concluding remarks. Following the conclusion are the bibliography and several appendices.

2. LITERATURE REVIEW PART I - ISSUES OF TASK MANAGEMENT

2.1 Flight Deck Automation Issues

In a general sense, today's commercial air carrier fleets are composed of two types of aircraft: advanced technology and traditional technology. The advanced technology aircraft incorporate a number of sophisticated automated systems that have the ability to perform tasks that in the past have been performed exclusively by the human pilots. These systems include such devices as the advanced autopilot, the Flight Management System (FMS), electronic instrument displays, and warning and alerting systems. Traditional technology aircraft are defined as lacking these types of automated systems. The presence of both types of aircraft in commercial fleets gives us a unique opportunity to compare them in present day operations.

2.1.1 Flight Management System

The Flight Management System (FMS) is one of the most novel devices on the advanced technology flight deck. The FMS is a highly integrated system that communicates with the pilot as well as other aircraft systems to perform a number of functions. Brief descriptions of some of these functions are presented in Table 2.1. The flightcrew interacts with the FMS via the Control and Display Unit (CDU) by entering information via the keypad and viewing the information displayed on the screen (see Figure 2.1).

Because of the FMS's complexity and relative novelty, it has received a lot of attention as a possible source for new errors. Recently, there have been a number of aviation human factors studies conducted that were specifically concerned with the effect the FMS has on the flightcrew's performance (Eldredge, Mangold, & Dodd, 1992; Sarter & Woods, 1992; Sarter, 1991).

TABLE 2.1 Brief description of some of the functions performed by the FMS.

Function	Description
Navigation	Determination of position, velocity and wind.
Performance	Trajectory determination, definition of guidance and control targets, flight path predictions. Time and fuel at destination.
Guidance	Error determination, steering and control command generation.
Electronic Instrument System	Computation of map and situation data for display.
Control and Display Unit (CDU)	Processing of keystrokes, flight plan construction, and presentation of performance and flight plan data.

Table adapted from Billings (1996).

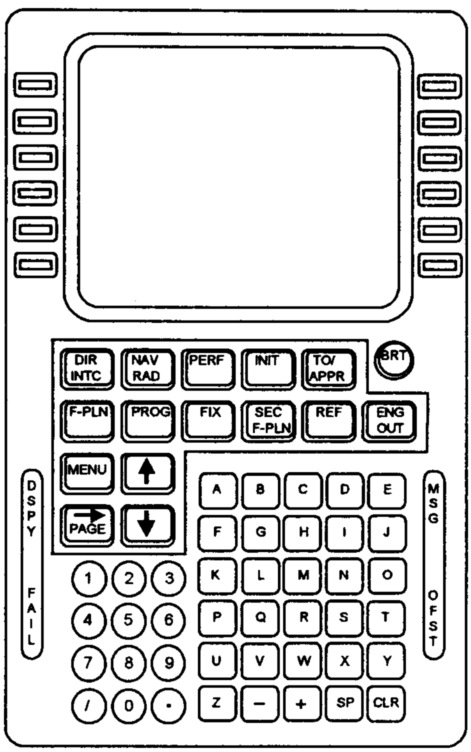


FIGURE 2.1 An example of a FMS CDU in an advanced technology aircraft.

2.1.2 Errors enabled by automation

While there is little doubt that technology has made significant contributions to both the safety and efficiency of operations, there are still concerns about replacing the human functioning with automated systems. With the introduction of automated systems, some flightcrew errors that had been a problem in the past have been significantly reduced (or eliminated). An example of this is the ability of the automated systems to track a precise heading with minimal deviations from the desired path in situations that pilots flying manually may err.

On the other hand, this functional change may also create the opportunity for errors that had not been possible in the past, or increase the chance of previously existing errors to occur; Wiener (1989) has referred to this as "enabling" errors. An example of a type of error that has been enabled is gross navigational deviations due to data entry errors. It was a gross navigational error of this type that contributed to the American Airlines flight 965 accident near Cali, Colombia. Aeronautica Civil, the aircraft accident investigating board of the Republic of Colombia, determined that during the approach into the airport in Cali, Colombia, the flightcrew "selected and executed a direct course to the identifier 'R,' in the mistaken belief that R was Rozo as it was identified on the approach chart. The pilots could not know without verification with the EHSI [Electronic Horizontal Situation Indicator] display or considerable calculation that instead of selecting Rozo, they had selected the Romeo beacon, located near Bogota, some 132 miles east-northeast of Cali" (Aeronautica Civil of the Republic of Colombia, 1997, p. 41). With the Romeo beacon programmed into the FMS, the airplane departed from its inbound course to Cali and flew east toward Bogota. When the flightcrew realized that they were off-course, they turned right to return to the extended centerline of the runway at Cali. At this point however, a direct course to the Cali airport led the aircraft into high mountainous terrain and shortly after their turn the aircraft impacted the side of a mountain. It would be highly unlikely that a gross navigational error such as this could occur without the automated systems.

Work focusing on these errors should not be to determine whether the traditional technology or the advanced technology produces the most errors overall, but rather to understand the errors that will be encountered in the new aircraft and how they may be successfully controlled.

At this time there is no comprehensive listing of what the errors enabled by the advanced technology flight deck are, but there are a number of ideas about what they may be. Funk et al. (in review) have compiled a list of 92 issues about automation on the flight deck from a broad range of sources including accident reports, incident report studies, surveys, and scientific experiments. To determine which of these issues should be valid concerns, they compiled a database of both supporting and contradictory evidence that addresses these issues. They found that many of the automation issues require further investigation to determine if they are indeed problems with which the aviation community should be concerned. This listing contains a diversity of issues including such things as problems arising because of either a lack of or excessive confidence that the pilots have in the automation, difficulties of transitioning between aircraft with different levels of automation, and the effect that the design of the automation has on the flightcrew's performance.

The theoretical concepts presented in the following two sections, Attention and Schema Theory, are related to human performance. These concepts will be used in a later section to hypothesize the relationship between automation on the commercial flight deck and Task Management (TM).

2.2 Attention

Some basic concepts of human cognitive processing capabilities and limitations are important to understanding the flightcrew's performance on the flight deck. The relevant psychology research done in the area of attention addressing this will be presented.

Though most people have an intuitive sense of what attention is, a good definition is difficult to find. Wickens (1984) has described the notion of attention as an "inferred underlying commodity, of limited availability, that enables performance of a task" (p. 67). In addition to the difficulty in finding a good definition, attention can be thought of both as the mechanism by which mental processing is concentrated (Kahneman, 1973), and as the concentrated mental processing itself (Eysenck & Keane, 1990). Here, I will refer to attention in the sense of the mental processing itself rather than as the mechanism.

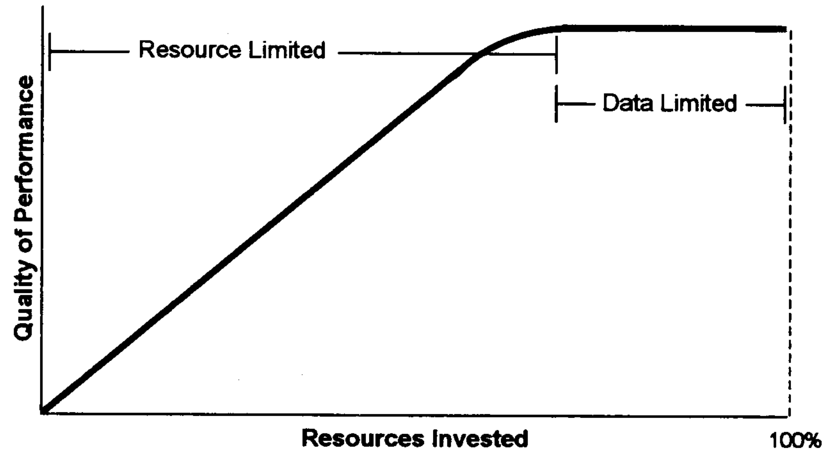
2.2.1 Dual-Task Paradigms

Much of the research that has been conducted to investigate attention has used the dual-task paradigm. The dual-task paradigm is the method in which a primary task and a secondary task are performed simultaneously. As the task parameters (i.e., task difficulty or instructions as to how much attention the subject should try to allocate to each task) are modified, the subject's performance on the tasks are measured. The subject's performance on the secondary task is then used as a gauge of how much attention the primary task is consuming.

One such study using the dual-task paradigm was conducted by North (1977). In this study, subjects were instructed to perform a digit processing task and a tracking task at the same time. The digit processing task difficulty was varied as the subject's performance on both tasks was measured. The results of this study were interesting in that North found that even though the task parameters of the primary task were modified, the performance of secondary task was unchanged.

2.2.2 Performance-Resource Function and Performance Operating Characteristic

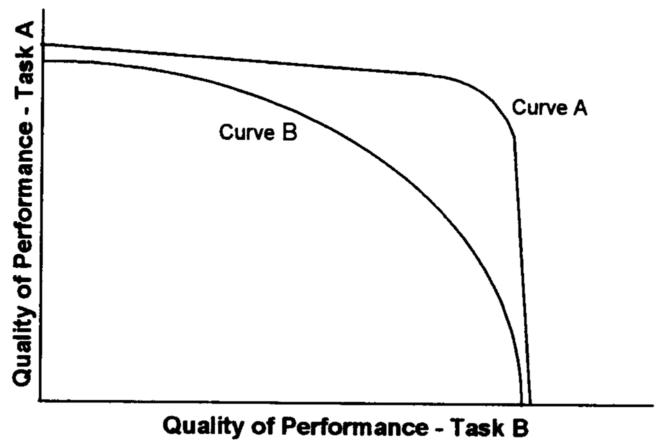
The Performance-Resource Function (PRF) is used to characterize how the amount of attention (i.e., resource) allocated to a task affects the quality of performance (Norman & Bobrow, 1975). The PRF can be used to illustrate two types of processing limitations that humans experience when performing a task. The first type of processing limitation is caused by the task being data-limited. A task is data-limited when it is not possible to improve the performance of a task beyond a certain point due to the quality of the data (see Figure 2.2). The second type of processing limitation is caused by the task being resource-limited. A task is considered resource-limited when the performance of the task changes as resources are added or taken away (see Figure 2.2).



Graph adapted from Wickens (1984).

FIGURE 2.2 Hypothetical Performance-Resource Function (PRF). Single task performance is at point S.

The Performance Operating Characteristic (POC) shown in Figure 2.3 has been developed to describe the extent to which two tasks can be performed together while the processing system is working at full capacity (Norman & Bobrow, 1975). The shape of the curve is determined by the amount of improvement that can be gained in one task by sacrificing the performance of the second task (Navon & Gopher, 1979).



Graph adapted from Wickens (1984).

FIGURE 2.3 Performance Operating Characteristic (POC). Curve A and Curve B each represents different pairs of tasks being performed simultaneously.

2.2.3 Single Resource Theory

Single Resource Theory posits that the human processing system is composed of one reservoir of undifferentiated resource (Wickens, 1984). It is able to describe our ability to perform multiple tasks simultaneously by asserting that attention can be allocated in graded quantity between separate activities. However, there are some problems with this view because it is unable to adequately account for some of the phenomena observed in dual-task studies.

Wickens (1984) has pointed out four such phenomena that Single Resource Theory cannot account for: difficulty insensitivity, perfect time-sharing, structural alteration effects, and uncoupling of difficulty and structure. Difficulty insensitivity refers to instances in which increases in demand or difficulty of a primary task can fail to influence the performance of a secondary task. Perfect time-sharing refers to instances when two moderately difficult tasks can be performed concurrently as well as they can be performed separately. Structural alteration effects refer to the situation in which the difficulty of the two tasks is held constant yet the amount of interference between the two tasks is increased by changing the processing structure (for example, changing the modality of a stimulus). Uncoupling of difficulty and structure refers to the situation in which two tasks of differing difficulty are each paired with a third task, and of these two tasks, it is the easier that interferes with the third task the most. While the first two of these phenomenon, difficulty insensitivity and perfect time-sharing, could be accounted for by assuming that the tasks involved possess very large data-limited regions, Wickens (1984) points out that it is doubtful that the tasks in the examples he has cited contained a large enough data-limited region to account for the phenomenon.

2.2.4 Multiple Resource Theory

The Multiple Resource Theory postulates that "there is more than one commodity within the human processing system that may be assigned resource-like properties (allocation, flexibility, sharing)" (Wickens, 1984, p. 78). Wickens (1980) has suggested that resources within the human processing system can be defined by a three-dimensional metric (see Figure 2.4). The first type of resource is defined by the way the processing is coded. This means that processing that is coded spatially requires a different resource than processing that is coded verbally. The second type of resource is defined by the processing stage. Wickens proposes that the perceptual encoding and central processing stages require resources that are functionally separate from the resources required by responding stage of processing. The third type of resource is defined by the stimulus and response modality. For each modality, such as visual and auditory, a functionally separate resource is required. Wickens points out that this remains the case even when peripheral interference (i.e., the physical constraints on processing) can be eliminated.

The Multiple Resource Theory, like the Single Resource Theory, is able to account for the human's ability to perform multiple tasks simultaneously. In addition, Wickens (1984) has shown that it can also account for the four phenomena described above for which the Single Resource Theory could not. Difficulty insensitivity can be explained by assuming that the resources required by the two tasks do not overlap. In this way, the resources used by the primary task would not affect the capacity of resources available to the secondary task and therefore the difficulty of the first task would have no effect on the performance of the secondary task. This same argument applies to the problem of perfect time-sharing. Perfect time-sharing would be possible when the resources required for each task to be performed did not overlap.

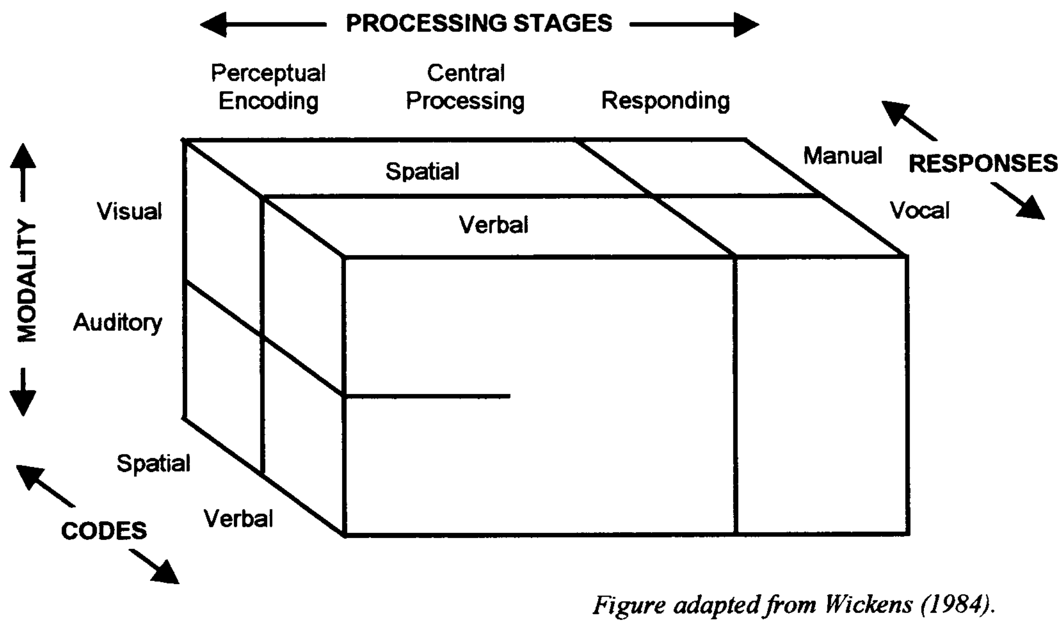


FIGURE 2.4 Three-dimensional metric proposed by Wickens (1984) that defines the resources within the human processing system.

Structural alteration effects would be expected due to the nature of Multiple Resource Theory. The processing structure is made up of several reservoirs, each having a maximum capacity. By structuring the tasks so that they both use overlapping processing resources, the interference between the tasks would occur once the capacity of the shared resources was exceeded.

Uncoupling of difficulty and structure can be explained by assuming that the easier of the two tasks and the third task each rely on the same resources while the more difficult of the two tasks and the third task rely on separate resources. In this way the easier and third task would generate more interference between one another than the more difficult task that is not competing for the same resource.

2.3 Schema Theory

Schema Theory is important in understanding human performance and Task Management (TM) in two ways. First, Schema Theory suggests a method for how the flightcrew may assess the state of the system. The assessment of the system allows the flightcrew to know what tasks must be prioritized and the attributes of the tasks necessary for prioritizing the relative importance among multiple tasks. Second, it is able to provide insight into the reason why attention may be inappropriately drawn toward some tasks. A general description of the theory is presented here.

"A schema is a structured cluster of concepts" (Eysenck & Keane, 1990, p. 275) that consists of relations between variable slots. The variable slots may be filled with concepts or other schemata, while the relations describe how the slots are connected. This notion can also be described in the language of artificial intelligence as a frame containing nodes and links (Winston, 1992). Schema represent generic knowledge that can be applied to a specific situation. "Within schema theory, the process of interpretation is guided by the principle that every perceived event must be mapped against some schema, and that all aspects of the schema must be compatible with the perceived information" (Adams, Tenny, & Pew, 1991, p. 17). This results in two types of processing. First, the data-driven, bottom-up type of processing occurs in which the system (e.g., a human) tries to fit a schema to the incoming stimuli. In this process the system attempts to explain the data by activating the appropriate schema. Second, the schema/hypothesis-driven, top-down processing occurs as the system looks for data to fill the slots of the currently active schema. This type of processing allows the system to anticipate the type of data that it could receive which leads to directing "attention and exploratory movements to particular aspects of the environment" (Adams, Tenny, & Pew, 1991, p. 18). It is thought that both types of processing may occur simultaneously to assure that the currently active schema is the most appropriate schema to be using (Adams, Tenny, & Pew, 1991).

Yates developed the ideas related to schema further by focusing on the content of awareness. Yates (1985) has argued that awareness is the result of a process of hypothesis generation based on one's past experiences to find an explanation for the incoming stimuli. A proposition that follows from Yates' (1985) work is "that attention will naturally be drawn to aspects of a scene that defy explanation within their currently active model of the environment" (Adams, Tenny, & Pew, 1991, p. 20).

2.4 Task Management

The general concepts of attention allocation and schema activation have laid the foundation to understanding the applied concept of Task Management (TM). But, it is important to bear in mind that the theories described above have been based primarily on studies in which the tasks to be performed overlap completely in time and are processed in parallel with each task allocated a share of attention. On the commercial flight deck, however, many of the tasks that require the flightcrew's attention cannot be performed simultaneously so the flightcrew must determine the order in which they will perform the tasks. The concept of TM has been developed to address the issue of allocating attention in such an environment (Adams, Tenny, & Pew, 1991).

TM refers to the function in which the human operator manages his/her available sensory and mental resources in a dynamic, complex, safety-critical environment in order to accomplish the multiple tasks competing for a limited quantity of attention. This function includes task initiation, monitoring, task prioritization, resource allocation, and task termination (Funk, 1991). Flightcrews must perform TM on the commercial flight deck because they do not possess the necessary resources to simultaneously execute all the tasks that demand their attention. The flightcrew must therefore prioritize the tasks in the order of most to least important and then allocate their resources according to this prioritization. In a dynamic system, the state of each task demanding attention continuously changes and as this occurs so too may change the relative urgency with which each task must be completed. Thus, the flightcrew must continuously perform the function of TM in order to maintain awareness of the changes in the state of the system and make the necessary revisions to the task prioritization.

While recently there have been several studies that have begun to look at TM on the commercial flight deck (Latorella, 1996; Rogers, 1996; Chou, Madhaven, & Funk, 1995), none of these studies has specifically addressed the relationship between TM and automation. There has been speculation that the level of automation on the flight deck may affect TM, in fact, one of the 92 automation issues identified by Funk et al. (in review) mentioned earlier concerns TM:

> issue167: The use of automation may make task management more difficult for flightcrews, possibly leading to unsafe conditions.

The reasons behind this speculation will be covered in a later section of this chapter.

2.5 Task Prioritization Errors

TM can be investigated by looking at the errors that flightcrews commit in prioritizing the tasks demanding attention. To do this it must be assumed that there is a "right" way and a "wrong" way to prioritize and that the ultimate prioritization of a flightcrew can be (at least partially) determined by observing the choice of tasks performed. Because the flightcrew is limited by the quantity of attention that they have available to distribute across the tasks they perform, they must manage tasks in such a way that higher priority tasks are allocated the available attention before lower priority tasks. If the flightcrew does not allocate his/her attention in this way, it is said that a Task Prioritization error is committed. Specifically, a Task Prioritization error is when the flightcrew gives their attention to a lower priority task to the detriment of a higher priority task. As the TM becomes more difficult, it is expected that the frequency of Task Prioritization errors would increase.

The prioritization strategy taught to every novice pilot is aviate, navigate, communicate, and manage systems. The tasks in the aviate category are concerned with using the flight systems and controls to fly the plane. The tasks in the navigate category are those concerned with planning the route and high-level route changes. The tasks in the communicate category are concerned with explicit communication with systems on the ground, between the flightcrew, with the cabin crew, with the company, and with the passengers. The tasks in the manage systems category are those concerned with assuring that the systems are operating normally and are capable of performing the functions necessary for the aviate tasks. Intuitively this rule of thumb makes sense. For example, it is easy to agree with the idea that keeping the plane in the air (i.e., aviate) is more important than making sure it is headed in the desired direction (i.e., navigate).

2.5.1 Attention and Task Prioritization Errors

The nature of TM, and therefore Task Prioritization errors, is related directly to attention. In TM, the tasks are prioritized and then resources are allocated to the tasks according to that plan. This allocation of resources would include attention allocation

By thinking of Task Prioritization in this way, it becomes apparent that it is related to Funk et al. (in review) issue102:

> issue102: The attentional demands of pilot-automation interaction may significantly interfere with performance of safety-critical tasks.

The relationship between this issue and Task Prioritization adds credence to the notion that Task Prioritization errors may present problems with which the aviation community should be concerned and therefore warrants further investigation.

2.5.2 Consequences of Task Prioritization Errors

Task prioritization errors can have disastrous consequences as evidenced by several accidents at least partially attributed to Task Prioritization errors. The following are two accidents in which the accident investigating board determined that misprioritization played a key role in the accident.[2]

---

[2] Accident investigations are highly detailed and intensive analyses conducted by experts concerning the circumstances involved in an aircraft accident. The final product of an investigation is an accident report in which the investigating board presents a set of findings, and the conclusions they have drawn, including the probable cause of the accident and other factors that may have contributed.

The first example is the often cited L-1011 Florida Everglades accident. On December 29, 1972, an Eastern Air Lines Lockheed L-1011 aircraft crashed approximately 18 miles west-northwest of Miami International Airport destroying the aircraft and killing 99 people on board. The National Transportation Safety Board (NTSB) determined that the probable cause of this accident was the flightcrew's failure to monitor the flight instruments during the final 4 minutes of the flight. Preoccupation with a malfunctioning nose landing gear position indicating system distracted the flightcrew's attention away from the instruments and allowed the descent to go unnoticed (NTSB, 1973). This would be considered a Task Prioritization error because the lower priority task of troubleshooting the malfunctioning landing gear indication (i.e., a manage systems task) was allocated attention while the higher priority task of maintaining the aircraft's altitude (i.e., an aviate task) was not allocated appropriate attention.

The second example is the Indian Airlines A320 accident in Bangalore, India. On April 14, 1990, an Indian Airlines Airbus A320 aircraft crashed just short of the runway at the Bangalore airport destroying the aircraft and killing 90 people on board. The investigators determined that the probable cause of the accident was the failure of the pilots to realize the gravity of the situation and immediately apply thrust. The pilots spent the final seconds of the flight trying to understand why the plane was in idle/open descent mode rather than taking appropriate action to avoid impact with the ground (Ministry of Civil Aviation, India, 1990). Again, the flightcrew committed a Task Prioritization error by allocating attention to the lower priority task of trying to understand the reason the automation was in a particular mode (i.e., a manage systems task) while the higher priority task of correcting the aircraft's descent (i.e., an aviate task) was not allocated appropriate attention.

Both of these accidents illustrate the disastrous consequences Task Prioritization errors can have. The Lockheed L-1011 is a traditional technology aircraft while the Airbus A320 is an advanced technology aircraft. These two accidents were chosen to illustrate two things: one, Task Prioritization errors occur in both advanced and traditional technology aircraft types; and two, the consequences of a Task Prioritization error can be equally fatal regardless of the aircraft type.

2.6 Why Automation May Affect Task Prioritization

An increasing number of accidents and incidents can be attributed to TM errors (e.g. Chou, Madhaven, & Funk, 1995). It has been speculated that higher levels of automation may make TM more difficult for flightcrews. There are several reasons behind this speculation.

2.6.1 Increase in the number of tasks

As pointed out by Rogers (1996), "it is expected as systems become more intelligent and complex, and more tasks and automated resources must be managed, flightcrew TM load will increase" (p. 239). All the flight control tasks found in the traditional technology aircraft must still be performed in the advanced technology aircraft but, in addition to these tasks, there are now tasks associated with communicating with and managing the automation. Adding tasks to the queue of tasks demanding attention increases the demands on the flightcrew. While the automation provides additional external resources for the flightcrew to utilize, these resources must be managed which increases demands on the function of TM. Prioritizing the tasks becomes more time consuming and effortful as more tasks that demand attention are added to the queue.

2.6.2 Overloading the same resources

Multiple Resource Theory predicts that when tasks are spread over several resources a human's overall processing capacity is effectively increased. On the other hand, when tasks require overlapping resources the capacity of one particular resource must be shared; the capacity of an individual resource is much smaller than that of the overall resources available in the processing system. This comes into play for the advanced technology flight deck in at least two ways. First, some of the demands added by automation require the cognitive processing resources that are already taxed in the traditional technology aircraft. For example, in the advanced technology aircraft "there is more information to be gathered and processed to ascertain the state of the aircraft and its automation" (Billings, 1996). Because of this, more prioritization is required because more tasks are demanding the same resource. This may cause prioritization to be more difficult in advanced technology aircraft.

Second, although probably to a lesser extent than cognitive processing, the input modality may affect the Task Prioritization in advanced aircraft. The monitoring tasks themselves on the advanced technology aircraft have changed from what they are in the traditional technology aircraft. In the traditional technology aircraft, visual, auditory and tactile sensations each play a large role in delivering information to the pilot. However, in the advanced technology aircraft, much of the information is received through the visual modality. The affect may be that the visual modality is overloaded and the tasks demanding attention must be prioritized to greater extent in order to share the visual resource.

2.6.3 FMS draws attention away from other tasks

The complexity of the FMS has the notorious reputation of being an "attention sink" (Williams, Tham, & Wickens, 1992). When the FMS fails to behave as expected, the flightcrew's attention can be drawn away from the highest priority tasks required for flying the aircraft. Two factors contribute to the ability of the FMS to draw the flightcrew's attention. First, because of the nature of the FMS the flightcrew cannot proceed with any other tasks until they either satisfy its needs or they turn it off. If pilots have an incentive to keep the FMS on, then they must correct the problem before their attention can be turned elsewhere.

Second, when the FMS fails to behave as expected the flightcrew's attention is drawn toward it as suggested by schema theory. As the functioning of the FMS defies explanation within the currently active schema, attention will be directed toward finding a better fitting schema. This phenomenon is sometimes referred to as 'novel pop out' (Johnston, Hawley, Plewe, Elliot, & DeWitt, 1990).

2.7 Primary Research Objective

Based on previous work in flight deck automation and the issues involved in TM, there appears to be a relationship between these two things. The primary objective of this study is to begin evaluating the relationship between the level of automation on the flight deck and attention allocation of commercial airline pilots by determining how automation affects the frequency of Task Prioritization errors as reported in ASRS incident reports.

3. LITERATURE REVIEW PART II - ASRS INCIDENT REPORT STUDIES

3.1 ASRS Incident Reports

The Aviation Safety Reporting System (ASRS) was created as a means to collect reports of situations that compromise safety so that strategies to prevent these situations from becoming accidents could be created (Chappell, 1994). These reports are called "incident reports" and are submitted voluntarily by aviation operations personnel with first-hand knowledge of the situation. Each report contains basic information about the aircraft involved and a description in the author's own words of the situation they have participated in or witnessed in which they believe safety has been compromised. Once processed by the ASRS, analysts remove any information that could identify the reporter and add the anonymous report to the database.

An example of an ASRS incident report is given in Figure 3.1. The abbreviations used can make the report difficult to understand so Figure 3.2 presents a more readable translation of this example.

Accession Number: 92507

Synopsis
ACR MLG ALT DEVIATION EXCURSION FROM CLRNC ALT. REPORTER SAYS FMA CHANGED FLT MODE AND ALT SELECT BY ITSELF.

Narrative
THE F/O WAS FLYING THE ACFT. WE HAD BEEN ISSUED SEVERAL VECTORS AND TURNS BY ATC FOR FLOW CTL INTO CHICAGO O'HARE. I WAS ON THE P/A EXPLAINING THE ENRTE DELAY TO THE PAX WHEN I NOTICED THE FMA HAD CHANGED FROM "PERF CRUISE" TO "PERF DSCNT," AND THE ALT SELECT HAD CHANGED FROM 35000 TO 33000'. I ASKED THE F/O IF WE HAD BEEN CLRED TO FL330. HE SAID NO. THE ACFT ALT WAS 34600' WHEN I NOTICED THE PROB. THE DSCNT WAS STOPPED AT 34500'. I DON'T KNOW WHY THE AUTOPLT ENTERED A DSCNT MODE. AN ALT WARNING DIDN'T OCCUR BECAUSE THE ALT SELECT HAD CHANGED ALSO. I SUSPECT A PWR SURGE IN THE ELECTRICAL SYS MAY HAVE CAUSED THE PROB. I HAVE EXPERIENCED THIS PROB IN THE PAST WITH THE MLG FLT GUIDANCE SYS WHEN A HYD PUMP IS TURNED FROM LOW TO HIGH.

FIGURE 3.1 Synopsis and Narrative of ASRS Incident Report #92507.

Accession Number: 92507

Synopsis
A medium-large transport aircraft used by an air carrier committed an altitude deviation. The aircraft made an excursion from the clearance altitude. The reporter says that the Flight Mode Annunciator (FMA) changed flight mode and altitude select by itself.

Narrative
The first officer was flying the aircraft. We had been issued several vectors and turns by Air Traffic Control to control the flow of traffic into Chicago O'Hare International Airport. I was on the public address explaining the enroute delay to the passengers when I noticed the FMA had changed from "PERF CRUISE" to "PERF DSCNT," and the altitude select had changed from 35000 to 33000 feet. I asked the first officer if we had been cleared to a flight level of 33000 feet. He said no. The aircraft's altitude was 34600 feet when I noticed the problem. The descent was stopped at 34500 feet. I don't know why the autopilot entered a descent mode. An altitude warning didn't occur because the altitude select had changed also. I suspect a power surge in the electrical system may have caused the problem. I have experienced this problem in the past with the medium-large aircraft flight guidance system when a hydraulic pump is turned from low to high.

FIGURE 3.2 Readable translation of Synopsis and Narrative of ASRS Incident Report #92507.

3.1.1 Strengths of ASRS Incident Reports

One of the strengths of ASRS incident reports is that they provide researchers with a practical alternative to collecting data from actual line operations. The ASRS incident reports contain descriptions of real line operations by individuals involved in aviation operations. These descriptions can be used to learn more about the operations.

Supporting evidence for the idea that ASRS incident reports provide a good representation of actual problems on the commercial flight deck is the similarity of proportions of occurrences by phase of flight between ASRS incident reports and aircraft accidents. Figure 3.3 shows that the ordering of instances by phase of flight of both ASRS incident reports and accidents is the same: terminal, initial, and cruise. While the proportions between ASRS incident reports and accidents are similar, they are not the same. The cause of this disparity may be attributed to the ground proximity in cruise. In cruise, the aircraft is far enough from the ground to allow the flightcrew time to recover from an unsafe condition before it is able to escalate from an incident into an accident. This would account for the lesser proportion of aircraft accidents found in cruise.

ASRS Incident Reports by Flight Phase

Total number of ASRS incident reports = 21,550*

	Initial Phases			Cruise	Terminal Phases		
	Takeoff	Initial Climb	Climb		Descent	Approach	Landing
Number of Incidents	2139	1100	3619	3970	4584	3681	2457
% of Total	9.9%	5.1%	16.8%	18.4%	21.3%	17.1%	11.4%
TOTALS	**31.8%**			**18.4%**	**49.8%**		

*These data are based on the incident reports submitted by air carriers available in the ASRS Aeroknowledge CD-ROM database (DOS Version Release 96-1). The categories listed are the ones that correspond to those available in the accident data, thus this is not an exhaustive listing of all categories available in ASRS incident reports.

Accidents by Flight Phase

Total number of accidents = 576**

	Initial Phases			Cruise	Terminal Phases		
	Takeoff	Initial Climb	Climb		Descent	Approach	Landing
Number of Accidents	81	56	42	27	37	199	125
% of Total	14.3%	9.9%	7.4%	4.8%	6.5%	35.1%	22.0%
TOTALS	**31.6%**			**4.8%**	**63.7%**		

**These data are based on Boeing Commercial Airplane Group (1997). *Statistical summary of commercial jet airplane accidents, worldwide operations, 1959-1996,* and exclude load, taxiing, and unload which account for 1.7% of all accidents.

FIGURE 3.3 Comparison of incidents (ASRS) and accidents by flight phase. These data are based on commercial air carriers and exclude military and general aviation.

Diehl (1991) further supports the idea that incidents provide a good representation of actual problems on the commercial flight deck with his discussion of accident generation. He has shown that in the aviation environment the rate of occurrence of some types of incidents is proportional to their accident rate. As an example of this, Diehl (1991) discusses the findings of the TWA B727 Berryville, Virginia accident investigation (NTSB, 1975). During this investigation, it was found that before this accident occurred many other pilots had been involved in similar situations which "except for luck, better weather, or lower mountains, would have resulted in the same type of catastrophe that befell this hapless TWA crew" (Diehl, 1991, p. 99). The FAA took this finding very seriously. In response, they established an incident reporting system that was "intended to identify unsafe operating conditions in order that they can be corrected before an accident occurs" (NTSB, 1975, p. 40). This reporting system eventually evolved into ASRS (Diehl, 1991).

3.1.2 Limitations of ASRS Incident Reports

While ASRS incident reports are a very valuable resource, several limitations of this type of data must be considered when interpreting the results of an ASRS based study. First, the collection of ASRS incident reports is a nonrandom sample of errors occurring in aviation operations. One cannot necessarily assume that the collection of ASRS incident reports have the same characteristics as the population of all errors committed in aviation operations (Chappell, 1994).

Second, the ASRS incident reports reflect reporting biases. The reporters submitting the incident reports are influenced by external motivators. These motivators could affect the validity and/or choice of information presented in the report. A few examples of possible motivators are ensuring immunity from penalization for error, a personal agenda, and unintentional personal bias (Williams, Tham, & Wickens, 1992).

3.2 Incident Report Studies

3.2.1 Descriptive Incident Report Studies

In the past, the ASRS has been primarily used by the aviation research community to conduct descriptive studies (Chappell, 1994). These descriptive analyses have been very valuable in yielding important information about line operations. For example, Williams, Tham, & Wickens (1992) conducted an ASRS study in which they reviewed two areas, failures in TM and failures in geographic orientation. They analyzed 158 reports related to TM and 100 reports related to geographic orientation. The result was a list of factors that they believed played a contributory role in each of these areas.

Palmer, Hutchins, Ritter, & vanCleemput (1993) performed another ASRS study. In this study, they analyzed 50 reports from traditional technology aircraft and 50 reports from advanced technology aircraft using a one page coding form to allow analysts to describe the reports in terms of the factors that were present in the incident. From this analysis, they constructed a descriptive framework of the flight deck as a single information processing system and used it to describe the error-tolerant properties of the system and why breakdowns occur. This study resulted in a number of recommendations for correcting the problems that were identified.

3.2.2 Inferential Incident Report Studies

While descriptive studies are the most common way to use the ASRS incident reports, they are not the only way. Inferential analyses may be conducted by carefully constructing a research question and then choosing an appropriate statistical test. There are very few statistical tests that are appropriate because it generally cannot be assumed that the incident data are normally distributed. One inferential statistic that does not rely on such an assumption is the Chi Square (χ^2) test. The χ^2 test can be used to determine if two or more incident types differ in the proportions of reports falling into various classifications, but to do this both the incident types and their classifications must be mutually exclusive. This can be accomplished by carefully constructing a research question (Chappell, 1994).

3.3 Secondary Research Objective

A methodology that could produce statistically significant findings (albeit with several caveats regarding the ASRS population) would be valuable. The secondary objective of this study was to create a methodology that models an effective way to use ASRS incident report data in an inferential analysis.

4. METHODS

4.1 Introduction

The objectives of this study were met by carefully constructing a study to ensure that a fair comparison was made between the advanced and traditional technology populations. To accomplish this, representative data samples were drawn from an ASRS incident report database and analyzed using an analysis tool constructed specifically for this study.

4.2 Sample Size Determination

Two samples of ASRS incident reports were compared in this study to determine if level of automation on the commercial aircraft flight deck affected the frequency of Task Prioritization errors. The first sample was composed of 210 incident reports submitted by pilots flying advanced technology aircraft and the second sample was composed of 210 incident reports submitted by pilots flying traditional technology aircraft. In total, 420 incident reports were analyzed.

The possibility exists that the effect of the level of technology of the aircraft could be confounded with differences in experience level because the advanced aircraft are comparatively new to commercial air carriers' fleets. To help avoid this confounding effect, the two samples were divided into three sub-samples each made up of 70 reports submitted during a specified time period: 1988-1989, 1990-1991, and 1992-1993. These submission periods were based on the availability of incident reports with narratives in the CD-ROM database used.

The sample sizes were determined by performing a power analysis. In a power analysis, the significance level, the desired power of the statistical test, and the effect size is used to calculate the sample size appropriate to efficiently conducting a methodologically sound study. The significance level, α, is "the standard of proof that the phenomenon exists, or the risk of mistakenly rejecting the null hypothesis" (Cohen, 1988, p. 4). "The power of a statistical test is the probability that it will yield statistically significant results" (Cohen, 1988, p. 1). The effect size is "the degree to which the phenomenon is present in the population, or the degree to which the null hypothesis is false" (Cohen, 1988, p. 9). A statistical test becomes more powerful (i.e., able to detect a smaller effect size) as the sample size is increased.

The power analysis was conducted using the following values: power = 0.80[3], significance level of $\alpha = 0.05$, and the effect size index of $w = 0.20$. It was determined that a sample size of 196 incident reports was required to reject the null hypothesis, or in other words, conclude that there is a significant difference between the frequencies found in the two samples. Because each sample was to be divided into 3 sub-samples (196/3 = 65.333), the sample size was rounded up to 210 (210/3 = 70).

A second power analysis was performed to determine if the sub-sample size of 70 was adequate. With the power = 0.80, significance level of $\alpha = 0.05$, and the effect size index = 0.40, it was determined that a sub-sample size of 49 incident reports was required to reject the null hypothesis. Because 70 is greater than 49, the sub-sample size of 70 was determined to be adequate.

It should be noted that the two power analyses conducted each used a different effect size index. The effect size index for each of the power analyses was chosen specifically for the effect that was to be detected. For the two aircraft technology type samples, I wanted to detect the smallest effect size without the sample size becoming prohibitively large. If a difference between the frequency rates of Task Prioritization errors between the two technology types existed, I wanted to detect it. The effect size index of 0.20 (loosely referred to as a 'medium-small' effect) was chosen for these samples. For the submission period sub-samples, I was interested only in detecting an effect of submission period that was large enough to significantly confound the effect of aircraft technology. It was not necessary to detect as small an effect for the sub-samples as was required for the aircraft technology type samples. Thus, I chose to use an effect

---

[3] When conducting a power analysis, it is a convention to set the power at 0.80 (Cohen, 1988).

size index of 0.40 (loosely referred to as a 'medium-large' effect) for the submission period sub-samples.

4.3 Report Selection Criteria

The ASRS incident reports used in this study were collected using the ASRS Aeroknowledge CD-ROM database (DOS Version Release 96-1). Homogeneity between samples is very important for statistical comparison studies. In an effort to collect homogenous samples, the sample populations[4] were constrained so that the level of automation (i.e., aircraft technology type) and the submission period were the only two differences between the samples. For example, all the reports from both the advanced technology and the traditional technology samples were constrained to reports submitted by a member of the flightcrew flying a two-person commercial air carrier aircraft in which the aircraft was classified as a medium-large transport, large transport or widebody transport aircraft (see Table 4.1).

Another parameter that was held constant was phase of flight. Based on the fact that over half of all commercial hull loss accidents (Boeing, 1997) and that approximately 50% of incidents reported to ASRS by commercial air carrier pilots occur during the terminal phases of flight (see Figure 3.2 in previous chapter), these phases of flight were considered a good place to look for errors. Thus, all reports analyzed occurred during the descent or approach phase of flight.

All population parameters used to collect the incident report samples are shown in Table 4.1.

[4] The term population is used here to denote the population of reports that meet the parameters defined. This usage of the term should not be confused with the population of all ASRS incident reports, or the population of all errors committed by flightcrews.

TABLE 4.1(a) Common population parameters for incident report samples.

Parameter	**Value**	**Definition**
Personnel number	P1	P1 was the first person coded with a sequence number; denoted the reporter.
Personnel Role	FLIGHT_CREW	The reporter was part of a flightcrew during the incident.
Personnel Affiliation	AIR_CARRIER	The reporter was affiliated with an air carrier.
Aircraft Handle	A1	A1 was the first aircraft coded with a sequence number; in this case, the aircraft P1 was flying.
Aircraft Type	MED_LARGE_TRANSPORT *or* LARGE_TRANSPORT *or* WIDE_BODY	The gross takeoff weight range was: 60,001 to 150,000 lbs. *or* 150,001 to 300,000 lbs. *or* over 300,000 lbs.
Crew Size	2	There were two flightcrew members on the aircraft, excluding observers and check airmen.
Operator Organization	AIR_CARRIER	The aircraft organization's principle mode of operation was air carrier (i.e., airline operator).
Flight Phase	APPROACH *or* DESCENT	The aircraft was in the approach or descent phase of flight during reported occurrence.

TABLE 4.1(b) Differentiating values for Aircraft Technology parameter for incident report samples.

Aircraft Technology	**Value**	**Definition**
Advanced Technology	EFIS_OR_HUD *or* INTEGRATED_NAV	CRT, HUD, and other advanced displays were installed on aircraft *or* FMS/FMC and INS were installed on the aircraft.
Traditional Technology	NOT_ADVANCED	No advanced technology equipment was installed on the aircraft.

4.4 Report Selection Methodology

To ensure that the sample was representative of the population, the reports were collected from the database in the following way. First, the six populations (i.e., the two aircraft technology populations each divided into three submission periods) were compiled from the database based on the population parameters described above. Second, based on the total number of reports in each of the six populations 70 random numbers for each sample were generated to determine which of the reports would be included in the sample. This allowed the samples to be drawn randomly without replacement. Third, the appropriate reports were then tagged and downloaded into a word processing document. Fourth, all information related to the report except for the ASRS number, the synopsis, and the narrative was removed. This was done so that the analyst would be unable to use this information to identify the report during analysis. Any information in the synopsis or the narrative that identified the report was not removed because the deletions would have left the report incomplete.

4.5 Analysis Tool

An incident analysis form was developed specifically for use in this project. This form allowed the analyst to classify the ASRS incident reports as either containing a Task Prioritization error or not based on the description given in the narrative of the report. Using the form, the analyst identified the tasks that were being performed during the incident period reported. Prioritization was evaluated by identifying whether the active tasks were related to the task categories of aviate, navigate, communicate, manage systems, or non-flight related tasks. If a task of lower priority was active while a task of higher priority that required resources was not active, the report was classified as containing a Task Prioritization error. An example of a blank analysis form is provided in Figure 4.1.

The incident analysis form contained a listing of all tasks that must be performed during the descent and approach phases of flight (see Figure 4.1). The task listing used was based on a functional analysis of a generic commercial air transport mission (Alter & Regal, 1992). These tasks were organized into four categories and the priority of the task was determined by the category to which it belonged (where 1 is highest and 5 is lowest): (1) Aviate, (2) Navigate, (3) Communicate, (4) Manage Systems, and (5) Non-Flight Related. There was no priority hierarchy within a category; it was assumed that all tasks that fell in a particular category were of the same priority. Each listed task was defined not only as performing the task itself, but also as maintaining awareness of the task's

Incident Report Analysis Form

Accession #: ____________________

Synopsis: ____________________

Descent/Approach Tasks: (check appropriate boxes and include explanatory comments)

REPORTED TASKS				STATUS during CRITICAL PERIOD			
Explicitly Stated	Strongly Implied	TASK LISTING	ACTIVE TASKS during CRITICAL PERIOD	Unknown	Satisfactory	Unsatisfactory	RELATED EXCERPTS / COMMENTS
		1. AVIATE TASKS					
☐	☐	1.1 Control/monitor aircraft configuration	☐	☐	☐	☐	
☐	☐	1.2 Control/monitor attitude	☐	☐	☐	☐	
☐	☐	1.3 Control/monitor lateral profile	☐	☐	☐	☐	
☐	☐	1.4 Control/monitor speed	☐	☐	☐	☐	
☐	☐	1.5 Control/monitor vertical profile	☐	☐	☐	☐	
☐	☐	1.6 Maintain clearances and restrictions	☐	☐	☐	☐	
☐	☐	1.7 Maintain separation with traffic, terrain	☐	☐	☐	☐	
		2. NAVIGATE TASKS					
☐	☐	2.1 Determine mode of lat/lon navigation	☐	☐	☐	☐	
☐	☐	2.2 Maintain awareness of temporal profile	☐	☐	☐	☐	
☐	☐	2.3 Modify route for weather, traffic, hazards	☐	☐	☐	☐	
☐	☐	2.4 Plan approach	☐	☐	☐	☐	
☐	☐	2.5 Program route in FMS	☐	☐	☐	☐	
☐	☐	2.6 Set navigational radios	☐	☐	☐	☐	
		3. COMMUNICATE TASKS					
☐	☐	3.1 Communicate with ATC	☐	☐	☐	☐	
☐	☐	3.2 Communicate with cabin crew	☐	☐	☐	☐	
☐	☐	3.3 Communicate with company	☐	☐	☐	☐	
☐	☐	3.4 Communicate with flightcrew	☐	☐	☐	☐	
☐	☐	3.5 Communicate with passengers	☐	☐	☐	☐	
☐	☐	3.6 Tune communication radios	☐	☐	☐	☐	
☐	☐	3.7 Uplink/ downlink information	☐	☐	☐	☐	
☐	☐	3.8 Receive ATIS	☐	☐	☐	☐	
		4. MANAGE SYSTEM TASKS					
☐	☐	4.1 Manage/correct system faults	☐	☐	☐	☐	
☐	☐	4.2 Monitor aircraft subsystems	☐	☐	☐	☐	
		5. NON-FLIGHT RELATED TASKS					
☐	☐	5.1	☐	☐	☐	☐	

Critical Period: ____________________

Task Prioritization: Was a Task Prioritization error committed? (circle one) **YES** **NO**

If YES, list the tasks involved in the prioritization error: ____________________

FIGURE 4.1 Blank Incident Report Analysis Form.

status. For example, the task '1.5 Control/monitor vertical profile' included controlling the vertical profile either manually or using the autopilot and monitoring the status of the vertical profile.

To illustrate how the analysis form was filled out, consider incident report #92507 that was discussed in Chapter 3 (see Figure 3.1) and its corresponding analysis form (see Figure 4.2).

Associated with each task listed on the form were three sets of boxes that were marked to highlight the parameters that were considered in the analysis. When any of the boxes were marked for a given task, the analyst entered an excerpt or short summary upon which the judgment to mark the box had been based in the column called 'Related Excerpt/Comment.'

Starting on the left, the first set of boxes, 'Reported Tasks,' were used to indicate all of the tasks that were reported as being performed during the block of time described in the incident. This set of boxes was used to give a rough summary of all tasks that the reporter had described. The analyst marked the 'explicitly stated' box if the reporter specifically mentioned the task in the narrative of the report. For example, given the following statement from incident report #92507:

> "WE HAD BEEN ISSUED SEVERAL VECTORS AND TURNS BY ATC FOR FLOW CTL INTO CHICAGO O'HARE."

The analyst would mark the 'explicitly stated' box for the Task 3.1 'Communicate with ATC' and include the excerpt 'ISSUED SEVERAL VECTORS AND TURNS BY ATC.' Reading on from this statement, it is implied, though not explicitly stated, that the flightcrew began to carry out these requests given by the ATC.

> "WE HAD BEEN ISSUED SEVERAL VECTORS AND TURNS BY ATC FOR FLOW CTL INTO CHICAGO O'HARE. I WAS ON THE P/A EXPLAINING THE ENRTE DELAY TO THE PAX..."

Incident Report Analysis Form

Accession #: 92507

Synopsis: ACR MLG ALT DEVIATION EXCURSION FROM CLRNC ALT. REPORTER SAYS FMA CHANGED FLT MODE AND ALT SELECT BY ITSELF

Descent/Approach Tasks: (check appropriate boxes and include explanatory comments)

REPORTED TASKS				STATUS during CRITICAL PERIOD			
Explicitly Stated	Strongly Implied	TASK LISTING	ACTIVE TASKS during CRITICAL PERIOD	Unknown	Satisfactory	Unsatisfactory	RELATED EXCERPTS / COMMENTS
		1. Aviate Tasks					
☐	☐	1.1 Control/monitor aircraft configuration	☐	☐	☐	☐	
☐	☐	1.2 Control/monitor attitude	☐	☐	☐	☐	
☐	✓	1.3 Control/monitor lateral profile	✓	✓	☐	☐	"ISSUED SEVERAL VECTORS AND TURNS BY ATC"
☐	☐	1.4 Control/monitor speed	☐	☐	☐	☐	
✓	☐	1.5 Control/monitor vertical profile	✓	☐	☐	✓	"FMA HAD CHANGED ... THE ACFT ALT WAS 34600' "
☐	☐	1.6 Maintain clearances and restrictions	☐	☐	☐	✓	"DSCNT WAS STOPPED AT 34500' "
☐	☐	1.7 Maintain separation with traffic, terrain	☐	☐	☐	☐	
		2. Navigate Tasks					
☐	☐	2.1 Determine mode of lat/lon navigation	☐	☐	☐	☐	
☐	☐	2.2 Maintain awareness of temporal profile	☐	☐	☐	☐	
☐	☐	2.3 Modify route for weather, traffic, hazards	☐	☐	☐	☐	
☐	☐	2.4 Plan approach	☐	☐	☐	☐	
☐	☐	2.5 Program route in FMS	☐	☐	☐	☐	
☐	☐	2.6 Set navigational radios	☐	☐	☐	☐	
		3. Communicate Tasks					
✓	☐	3.1 Communicate with ATC	☑	☑	☐	☐	"ISSUED SEVERAL VECTORS AND TURNS BY ATC"
☐	☐	3.2 Communicate with cabin crew	☐	☐	☐	☐	
☐	☐	3.3 Communicate with company	☐	☐	☐	☐	
✓	☐	3.4 Communicate with flightcrew	✓	☐	✓	☐	"I ASKED F/O IF WE HAD BEEN CLEARED"
✓	☐	3.5 Communicate with passengers	✓	✓	☐	☐	"I WAS ON THE P/A"
☐	☐	3.6 Tune communication radios	☐	☐	☐	☐	
☐	☐	3.7 Uplink/ downlink information	☐	☐	☐	☐	
☐	☐	3.8 Receive ATIS	☐	☐	☐	☐	
		4. Manage System Tasks					
☐	☐	4.1 Manage/correct system faults	☐	☐	☐	☐	
☐	☐	4.2 Monitor aircraft subsystems	☐	☐	☐	☐	
		5. Non-Flight Related Tasks					
☐	☐	5.1	☐	☐	☐	☐	

Critical Period: "GIVEN CLEARANCE ALTITUDE" to "I NOTICED THE PROBLEM"

Task Prioritization: Was a Task Prioritization error committed? (circle one) (**YES**) **NO**

If YES, list the tasks involved in the prioritization error: 1.6, 3.4, 3.5

FIGURE 4.2 Incident Report Analysis Form completed for ASRS Incident Report #92507.

The analyst would mark the 'strongly implied' box for the Task 1.3 'Control/monitor lateral profile' and again include the excerpt 'ISSUED SEVERAL VECTORS AND TURNS BY ATC.'

The next box, 'ACTIVE TASKS during CRITICAL PERIOD,' was marked when the task was active during the critical period of the incident. The critical period consisted of all the events that took place between the time that the "desired state" was defined and the time that the flightcrew became aware that the desired state was not or would not be met (i.e., a deviation occurred). The analyst entered the critical period in the appropriate space at the bottom of the form. In incident report #92507, the critical period was "given clearance altitude" to "I noticed the problem." This would indicate that the critical period included all tasks that occurred between the point that the desired state of maintaining the cleared altitude was declared and the point that the flightcrew realized that they had overshot this altitude. In this report, the clearance for their desired altitude had been given before the window of time described in this incident report so all the tasks described up to the point that the captain noticed the problem were considered active tasks.

The last set of boxes, 'STATUS during CRITICAL PERIOD,' were marked if the task was active during the critical period (i.e., had been marked 'ACTIVE TASKS during CRITICAL PERIOD'). The 'Unknown' box was marked when the analyst was unable to discern the task's status from the narrative. For example, it cannot be determined from this narrative if the public address system was working correctly and that the passengers actually heard the captain's announcement. In this case the analyst would mark Task 3.5 'Communicate with passengers' as status 'Unknown.'

The 'Satisfactory' box was marked when the desired state of the task had and/or would be achieved given the current trend of activities. For example, given the following statement:

> "...I ASKED THE F/O IF WE HAD BEEN CLRED TO FL330. HE SAID NO..."

The analyst would mark the 'Satisfactory' box for the Task 3.4 'Communicate with flightcrew'. The first officer and the captain effectively communicated this information.

The 'Unsatisfactory' box was marked when the reporter stated in the narrative that the desired state of the task had not and/or would not be achieved given the current trend of activities. For example, given the following statement:

> "...THE ALT SELECT HAD CHANGED FROM 35000 TO 33000'. I ASKED THE F/O IF WE HAD BEEN CLRED TO FL330. HE SAID NO. THE ACFT ALT WAS 34600' WHEN I NOTICED THE PROB..."

The analyst would mark the 'Unsatisfactory' box for the Task 1.6 'Maintain clearances and restrictions.' In this example, the desired altitude was 35,000 feet yet the altitude of the aircraft was 34,600 feet, a discrepancy of 400 feet.

Once all the appropriate boxes were marked on the analysis form, the incident report was classified as to whether a Task Prioritization error was committed by circling 'yes' or 'no.' The classification was determined by using the following rule:

> If the status of a higher priority task is unsatisfactory and it is not active AND a lower priority task is active, then the incident report is classified as "TP error occurred" (otherwise it is classified as "no TP error occurred").

When a report was classified as containing a Task Prioritization error then the tasks involved in this error were listed in the space provided at the bottom of the analysis form. In incident report #92507, Task 1.6 'Maintain clearances and restrictions' was not active and unsatisfactory while the lower priority tasks 3.4 'Communicate with flightcrew' and 3.5 'Communicate with passengers' were active, thus this incident report was classified as containing a Task Prioritization error and these tasks were listed at the bottom of the form.

4.6 Application of the Analysis Tool

Each incident report was analyzed using the incident report analysis form described above. To minimize bias during the analysis, the two samples (including the three sub-samples within each) were randomly mixed and the sample to which each incident report belonged was not specified until all analyses were complete. After all reports had been analyzed, the reports were sorted and the data summarized.

4.7 Statistical Analysis

The objective of this study was to determine if there was a significant difference between the frequency of Task Prioritization errors in advanced technology and traditional technology aircraft, or as stated in terms of a null and alternate hypothesis:

> NULL HYPOTHESIS (**H0**): There is **no** significant difference between the number of incident reports from advanced technology aircraft in which a Task Prioritization error occurred and the number of incident reports from traditional technology aircraft in which a Task Prioritization error occurred.
> ALTERNATE HYPOTHESIS (**H1**): There is a significant difference.

The Chi Square (χ^2) test was chosen to test these hypotheses. This statistical test allowed me to determine if there was a real difference between the two samples and not simply a difference that could be attributed to sampling error. The χ^2 test has several requirements, and as Table 4.2 shows, all were met in this study.

The requirements of the statistical binomial test were closely met and thus the binomial test was also considered for use in this study. The binomial test is appropriate for studies with data in which there are only two possible outcomes, for example either containing a Task Prioritization error or not. This test computes exact probabilities and allows the analyst to perform directional tests. In this study however, it could not be used because the sample size was far too large. When using the binomial test, the computations become quite cumbersome for sample sizes greater than 25 (Sharp, 1979).

TABLE 4.2 Meeting χ^2 Test Requirements

χ^2 Test Requirements	How Requirements Were Met In This Study
Nominal Data	The data consisted of incident reports classified as either containing a Task Prioritization error or not.
One Or More Categories	There were two categories: classified as either containing a Task Prioritization error or not.
Independent Observations	Each incident report represented an independent observation.
Adequate Sample Size (i.e., Greater Than Five)	The sample size was 210.
Simple Random Sample	The samples were drawn randomly without replacement
Data In Frequency Form	The frequency of phenomenon was determined by counting the number of reports from each sample that were classified as containing a Task Prioritization error and the number of reports that were not classified as containing a Task Prioritization error.
All Observations Used	All the observations were used.
Two Tailed Test Only	A directional (i.e., two tailed) test was not required, in this study it was only important to determine whether or not a statistically significant difference existed.

Table adapted from Sharp (1979).

The next step in this study was to determine if the time period in which the report was submitted had an effect on the frequency of Task Prioritization errors. For the advanced technology aircraft sample, this was stated in terms of the following null and alternate hypotheses:

NULL HYPOTHESIS **(H0)**: There is **no** significant difference in the number of incident reports in which a Task Prioritization error occurred among the three sub-samples of the advanced technology aircraft sample.
ALTERNATE HYPOTHESIS **(H1)**: There is a significant difference.

These hypotheses were tested using the χ^2 test for the same reasons that it was used for the first set of hypotheses tested in this study. Similarly for the traditional technology sample, the χ^2 test was used to test the following null and alternate hypotheses:

NULL HYPOTHESIS **(H0)**: There is **no** significant difference in the number of incident reports in which a Task Prioritization error occurred among the three sub-samples of the traditional technology aircraft sample.
ALTERNATE HYPOTHESIS **(H1)**: There is a significant difference.

5. RESULTS

5.1 Advanced versus Traditional Technology

5.1.1 Overall Effect of Technology

Of the 420 incidents reports analyzed, 43 (10.2%) were classified as containing Task Prioritization errors. Of these, 28 were from the advanced technology sample and 15 were from the traditional technology sample (see Table 5.1).

TABLE 5.1 Summary of the frequencies of Task Prioritization errors.

	Task Prioritization Error Frequency		
Submission Period	**Advanced Technology**	**Traditional Technology**	**Total Errors by Submission Period**
1988-1989	13	7	20
1990-1991	11	5	16
1992-1993	4	3	7
Total Errors by Aircraft Technology	28	15	

The Chi Square (χ^2) test was used to determine if the difference between the 28 Task Prioritization errors found in advanced technology incident reports and the 15 Task Prioritization errors in traditional technology aircraft was statistically significant.

The χ^2 test was performed in three steps: first, the expected values were calculated; second, the χ^2 calculation itself was performed; and third, the χ^2 value was compared to a predetermined Critical Value. The first step in performing the χ^2 test was to determine the hypothetical frequency of the phenomenon that would be expected if the null hypothesis were true. In this study, the expected value was calculated based on the observed frequencies of Task Prioritization errors found in the data. This was found for each of the two samples by calculating the ratio of the product of row totals and column totals to the grand total. For example, consider the expected value for the number of advanced technology reports classified as containing a Task Prioritization error. This was found by multiplying the total number of reports classified as having a Task Prioritization error (i.e., 43) by the total number of reports in the advanced technology aircraft sample (i.e., 210) and dividing this number by the grand total (i.e., 420). These calculations are shown in Table 5.2.

TABLE 5.2 Calculating the expected values from the observed frequencies of Task Prioritization errors.

	Observed Frequencies		
Report Type	**Error**	**No Error**	*row totals*
Advanced Technology	28	182	*210*
Traditional Technology	15	195	*210*
column totals	*43*	*377*	

Grand total = 420

	Expected Values	
Report Type	**Error**	**No Error**
Advanced Technology	$\frac{43 \times 210}{420} = 21.5$	$\frac{377 \times 210}{210} = 188.5$
Traditional Technology	$\frac{43 \times 210}{420} = 21.5$	$\frac{377 \times 210}{210} = 188.5$

After the expected values were found, the χ^2 value was calculated using the following equation and substituting in the appropriate values:

$$\chi^2 = \sum \frac{(\text{Observed Frequency} - \text{Expected Value})^2}{\text{Expected Value}}$$

$$\chi^2 = \frac{(28-21.5)^2}{21.5} + \frac{(15-21.5)^2}{21.5} + \frac{(182-188.5)^2}{188.5} + \frac{(195-188.5)^2}{188.5}$$

$$\chi^2 = \underline{\mathbf{4.379}}$$

The final step is to compare the χ^2 value found in the previous step to the Critical Value. The Critical Value is "the value of a test statistic at or beyond which we will reject H_0 [Null Hypothesis]" (Howell, 1989, p. 150). The Critical Value is based on the degrees of freedom of the test and the significance level that has been selected by the analyst. The degrees of freedom (*df*) was found using the following equation:

Degrees of Freedom = *df* = (Number of Rows – 1) x (Number of Columns – 1)

df = (2 – 1) x (2 – 1)
df = 1

A significance level of $\alpha = 0.05$ was selected. At this significance level the null hypothesis can be rejected with a 95% confidence level or, alternatively stated, there is only a 5% chance that a Type I error (i.e., rejecting the null hypothesis when it is true) would be made.

Looking in a χ^2 Table of Critical Values it is found that at 1 degree of freedom and a significance level of $\alpha = 0.05$, the Critical Value is 3.84 (Howell, 1989, p. 329). Because the χ^2 value of 4.379 is greater than the Critical Value of 3.84, the null hypothesis can be rejected with a 95% confidence level.

A Microsoft Excel function called "CHITEST" was used as a secondary means determine significance with the χ^2 test. CHITEST returns the p-value of the test. The p-value is the precise level at which the null hypothesis can be rejected. CHITEST returned 0.036 meaning that the null hypothesis can be rejected with a 96.4% confidence level. This result corroborates the hand calculations.

5.1.2 Effect of Technology by Submission Period

The χ^2 test was used next to compare the frequency difference between advanced technology and traditional technology aircraft by submission period. For each of the three submission periods, the difference between the technology types was not statistically significant (p-value > 0.10).

5.2 Incident Report Submission Period

The two samples were divided into three sub-samples each made up of 70 reports submitted during a specified time period: 1988-1989, 1990-1991, and 1992-1993. The Task Prioritization error frequencies by submission period were summarized in Table 5.1.

5.2.1 Overall Effect of Submission Period

The data for each submission period from both the advanced technology and the traditional technology aircraft were combined. The χ^2 test was used to determine if the differences between the submission periods were significant. The χ^2 value was 6.891 at 2 degrees of freedom with a p-value of 0.032. Using a significance level of $\alpha = 0.05$, it was concluded that the null hypothesis could be rejected and thus this difference was statistically significant.

Further χ^2 tests to compare the difference between pairs of submission dates were conducted using the combined data from both the advanced technology and traditional technology aircraft samples. The results of these tests are given in Table 5.3.

TABLE 5.3 Results of χ^2 tests for comparing pairs of submission dates using data combined from both the advanced technology and traditional technology aircraft samples.

Submission Periods Tested	Chi Square Value	Degrees of Freedom	p-Value
1988-1989 vs. 1990-1991	0.510	1	0.475
1990-1991 vs. 1992-1993	3.837	1	0.050
1988-1989 vs. 1992-1993	6.927	1	0.008

5.2.2 Effect of Submission Period on Advanced Technology Sample

The data from the advanced technology aircraft only was used, and the χ^2 value was calculated to compare the three submission periods. The χ^2 value was 5.522 at 2 degrees of freedom with a p-value of 0.063. This would be significant at $\alpha = 0.10$.

Additional χ^2 tests were also conducted using the advanced technology sample data to compare the differences between pairs of submission dates (see Table 5.4).

TABLE 5.4 Results of χ^2 tests for comparing pairs of submission dates using data from the advanced technology aircraft sample.

Submission Periods Tested	Chi Square Value	Degrees of Freedom	p-Value
1988-1989 vs. 1990-1991	0.201	1	0.654
1990-1991 vs. 1992-1993	3.658	1	0.056
1988-1989 vs. 1992-1993	5.423	1	0.020

5.2.3 Effect of Submission Period on Traditional Technology Sample

The same approach taken in analyzing the advanced technology sample frequency data by submission period was used to analyze the traditional technology data. The result was not statistically significant (p-value = 0.423).

5.3 Tasks Involved

Further analysis was conducted to determine if a pattern could be found related to the tasks involved in the Task Prioritization errors found. To do this, the tasks involved in the reports with Task Prioritization errors were summarized by computing the raw number of times a task was involved in a Task Prioritization error (see Table 5.5).

Next, the tasks with the largest raw difference between the number of times they were involved in Task Prioritization errors in advanced technology aircraft reports and the traditional technology reports were identified. The tasks found to have the greatest difference (excluding Task 2.5 "Program route in FMS" which could not occur in the traditional technology aircraft) were Task 3.2 "Communicate with cabin crew" and Task 3.3 "Communicate with company." For these tasks, the χ^2 test was used to determine if these differences were statistically significant. For both tasks, it was found that the difference was not statistically significant (respectively, p-value = 0.370 and p-value = 0.313).

TABLE 5.5 Summary of the tasks involved in Task Prioritization errors for advanced technology and traditional technology samples. Asterisks (*) have been placed next to the tasks with the greatest difference between occurrences in the two aircraft technology types.

Tasks	Involvement Frequency: Advanced Technology	Involvement Frequency: Traditional Technology	
2.1 Determine mode of lat/lon navigation	0	0	
2.2 Maintain awareness of temporal profile	0	0	
2.3 Modify route for weather, traffic, hazards	1	0	
2.4 Plan approach	4	2	
2.5 Program route in FMS	12	0	*
2.6 Set navigational radios	1	1	
3.1 Communicate with ATC	5	2	
3.2 Communicate with cabin crew	7	2	*
3.3 Communicate with company	5	1	*
3.4 Communicate with flightcrew	8	7	
3.5 Communicate with passengers	2	0	
3.6 Tune communication radios	2	0	
3.7 Uplink/ downlink information	0	0	
3.8 Receive ATIS	1	3	
4.1 Manage/correct system faults	2	1	
4.2 Monitor aircraft subsystems	2	3	
5.1 Non-flight Related Tasks *(various)*	3	1	

6. DISCUSSION

6.1 Primary Objective

The primary objective of this study was to begin evaluating the relationship between TM of commercial airline pilots and the level of automation on the flight deck by determining how automation affects the frequency of Task Prioritization errors as reported in ASRS incident reports. I found that Task Prioritization errors occurred in both advanced technology and traditional technology aircraft, and that overall there was a statistically significant difference between the number of reports classified as containing a Task Prioritization error in the advanced and traditional technology aircraft. This difference in the frequency of Task Prioritization errors suggests that TM may be more difficult in the advanced technology aircraft.

I cannot unequivocally state that the difference was caused by the nature of the design of the automation because this is confounded by the novelty of the advanced aircraft in air carrier fleets. In an attempt to better understand the effect of aircraft technology type, I looked more closely at the difference by submission period between the advanced technology and traditional technology samples. However, I found that the difference by submission period between aircraft technology types was not statistically significant. Why would this be the case? I believe the answer lies in the power of the statistical test. For the overall test in which the three submission periods' frequency data were combined for the two technology types, the power of the test was such that a medium-small effect could be detected. For the tests conducted by submission period, however, the power of the test was such that a medium-large effect could be detected. This difference in effect size detection was due to the difference in sample size. In the population of ASRS incident reports, the actual effect that I was trying to detect was smaller than medium-large and therefore the test by submission period lacked the appropriate power to detect it. To determine if there was a significant difference between aircraft technology in each submission period, the sub-sample size would have to be increased.

I also looked at the effect of submission period on Task Prioritization errors. By separating the two samples into three equal sub-samples based on submission period, a decrease in the frequency of Task Prioritization errors in both the advanced technology sample and the traditional technology sample over time became apparent. This difference was statistically significant for the advanced technology sample; however, it was not statistically significant for the traditional technology sample. These data are consistent with the idea that industry experience with the advanced technology aircraft played a role in the differences in the frequency of Task Prioritization errors, but this cannot be stated conclusively. It may be the case that improved pilot training programs, or any number of other factors could have contributed to this reduction in Task Prioritization errors and that this reduction may have occurred in all aircraft, regardless of their level of technology. Further research is required to determine if the novelty of the advanced aircraft indeed played the critical role in creating the difference of frequency in Task Prioritization errors between the two aircraft technology types.

When evaluating the results of this study, one must bear in mind the limitations of ASRS incident report data. The samples used in this study were drawn from a nonrandom sample of events occurring in aviation operations and the ASRS incident reports themselves reflect reporting biases. What can be said with confidence, however, is that Task Prioritization errors do exist in actual line operations and their existence warrants thoughtful consideration. This study shed some light on one factor, automation on the commercial flight deck, which may effect the frequency of these errors.

6.2 Secondary Objective

The secondary objective of this study was to create an effective methodology for using ASRS incident reports for inferential analysis. By carefully constructing a research question and choosing an appropriate statistical test, an inferential analysis was conducted on the data collected. In this study statistically significant results were derived, supporting the notion that ASRS incident reports can be effectively used both for descriptive analyses and for inferential analyses.

By using ASRS data, I took advantage of several of the strengths of this type of data. First, the reports were able to provide a practical alternative to collecting data from the jumpseat of a commercial aircraft. The situations described in the narratives of the reports represented situations that had occurred in line operations which gave this study ecological validity and avoided the possibility that the effect found was an artifact of a laboratory experiment. Second, the large number of incident reports available made it possible to construct a study with a large enough power to detect a medium-small effect.

6.3 Conclusion and Recommendations

While Task Prioritization errors occur in both advanced technology and traditional technology aircraft, these errors occur more frequently in the advanced technology aircraft. The increased frequency of Task Prioritization errors suggests that TM may be more difficult in advanced technology aircraft. The submission period effect suggests that there is a downward trend in Task Prioritization errors in advanced technology aircraft.

Based on these conclusions, there are two recommendations that I would like to make. First, I recommend that further research be conducted to differentiate the effect of automation due to the nature of its design and the effect of automation based on its novelty in air carrier fleets. One way this could be accomplished is by analyzing additional submission periods and adding these data to the results presented here. The results of such a study could also be used to determine if the overall downward trend of Task Prioritization errors that appeared in this study continues.

Second, I recommend that when designing a training program for pilots of advanced technology aircraft that this information be disseminated to the pilots. The information could raise the awareness of pilot's susceptibility to Task Prioritization errors in advanced technology aircraft. It is possible that a heightened awareness could counteract this susceptibility.

BIBLIOGRAPHY

Adams, M.J., Tenny, Y.J., & Pew, R.W. (1991). State-of-the-Art Report: Strategic workload and the cognitive management of advanced multi-task systems. SOAR CSERIAC 91-6. Wright-Patterson Air Force Base, OH: Crew System Ergonomics Information Analysis Center.

Aeronautica Civil Of The Republic Of Colombia (1997). *Aircraft Accident Report: Controlled flight into terrain, American Airlines flight 965, B757-223, near Cali, Colombia, December 20, 1995.*

Alter, K.W., & Regal, D.M. (1992). *Definition of the 2005 flight deck environment.* NASA Contract Report 4479. Washington, DC: National Aeronautics and Space Administration.

Billings, C.E. (1996). *Human-centered aircraft automation: Principles and guidelines.* NASA Technical Memorandum 110381. Moffett Field, California: NASA-Ames Research Center.

Boeing Commercial Airplane Group (1997). *Statistical summary of commercial jet airplane accidents, worldwide operations, 1959-1996.* Seattle, WA: Boeing Commercial Airplane Group, Airplane Safety Engineering.

Broadbent, D.E. (1982). Task combination and selective intake of information. *Acta Psychologica, 50,* pp. 253-290.

Chappell, S.L. (1994). Using voluntary incident reports for human factors evaluations. In N. Johnston, N. McDonald, & R. Fuller (Eds.), *Aviation Psychology in Practice.* Brookfield, VT: Avebury Technical, pp. 149-169.

Cherry, E.C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America, 25,* pp. 975-979.

Chou, C., Madhaven, D., & Funk, K. (1995). Studies of cockpit task management errors. *International Journal of Aviation Psychology 6*(4), pp. 307-320.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Diehl, A.E. (1991). Human performance and systems safety considerations in aviation mishaps. *International Journal of Aviation Psychology 1*(2), pp. 97-106.

Eldredge, D., Mangold, S., & Dodd, R.S. (1992). *A review and discussion of flight management systems incidents reported to the Aviation Safety Reporting System*. Final Report No. DOT-FAA-RD-92/2. Washington, D.C: Federal Aviation Administration.

Eysenck, M.W., & Keane, M.T. (1990). *Cognitive psychology: A student's handbook*. London: Lawrence Erlbaum Associates, Ltd.

Funk, K., Lyall, B., Wilson, J., Niemczyk, M., Vint, R., Suroteguh, C., & Owen, G. (in review). Flight deck automation issues. *International Journal of Aviation Psychology*.

Howell, D.C. (1989). *Fundamental statistics for the behavioral sciences* (2nd ed.). Boston, MA: PWS-KENT Publishing Company.

Johnston, W.A., Hawley, K.J., Plewe, S.H., Elliot, J.M.G., & DeWitt, M.J. (1990). Attention capture by novel stimuli . *Journal of Experimental Psychology: General, 119*, pp. 397-411.

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

Latorella, K.A. (1996). Investigating interruptions: An example from the flightdeck. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* held in Philadelphia, pp. 249-253.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86*, pp. 214-255.

Norman, D.A., & Bobrow, D.J. (1975). On data-limited and resource-limited processes. *Coginitve Psychology*, 7, pp. 44-64.

North, R.A. (1977). *Task components and demands as factors in dual-task performance*. Report No. ARL-77-2/AFOSE-77-2. Urbana, IL: University of Illinois at Urbana-Champaign, Aviation Research Laboratory.

NTSB. (1973). *Aircraft accident report. Eastern Air Lines, Incorporated, L-1011, N310EA, Miami, Florida, December 29, 1972.* Report No. NTSB-AAR-73-14. Washington, DC: National Transportation Safety Board.

NTSB. (1975). *Aircraft accident report. Trans World Airlines, Incorporated, Boeing 727-231, N54328, Berryville, Virginia, December 1, 1974.* (Report No. NTSB-AAR-75-16). Washington, DC: National Transportation Safety Board.

Ministry of Civil Aviation, India. (1990). *Report on accident to Indian Airlines Airbus A-320 Aircraft VT-EPN at Bangalore on 14th February 1990.* Government of India, Ministry of Civil Aviation.

Palmer, E.A., Hutchins, E.L., Ritter, R.D., & VanCleemput, I. (1993). *Altitude deviations: Breakdown of an error-tolerant system*. NASA Technical Memorandum 108788. Moffett Field, California: NASA-Ames Research Center.

Rogers, W.H. (1996). Flight deck task management: A cognitive engineering analysis. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* held in Philadelphia, pp. 239-243.

Sarter, N.B. (1991). The Flight Management System - pilots' interaction with cockpit automation. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics.* New York: IEEE, pp. 1307-1310.

Sarter, N.B., & Woods, D.D. (1992). Pilot interaction with cockpit automation: operational experiences with the Flight Management System. *International Journal of Aviation Psychology* 2(4), pp. 303-321.

Sharp, V.F. (1979). *Statistics for the social sciences.* Boston, MA: Little, Brown and Company.

Schutte, P.C., & Trujillo, A.C. (1996). Flight crew task management in non-normal situations. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* held in Philadelphia, pp. 244-248.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and Performance VIII.* Hillsdale, N.J.: Erlbaum, pp. 239-257.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & D.A. Davies (Eds.), *Varieties of Attention.* Orlando, FL: Academic Press, pp. 63-102.

Wiener, E.L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft.* NASA Contract Report 177528. Moffett Field, California: NASA-Ames Research Center.

Wiener, E.L., & Curry, R.E. (1980). Flight deck automation: Promises and problems. *Ergonomics, 23,* pp. 995-1011.

Williams, H.P., Tham, M., & Wickens, C.D. (1992). *Resource management and geographic disorientation in aviation incidents: A review of the ASRS data base.* NASA Technical Report ARL-92-3/NASA-92-2. Moffett Field, CA: NASA Ames Research Center.

Winston, P.H. (1992). *Artificial intelligence* (3rd ed.). Reading, MA: Addison-Wesley Publishing Company.

Yates, J. (1985). The content of awareness is a model of the world. *Psychological Review, 92,* pp. 249-284.

APPENDICES

Appendix A: Glossary

The glossary includes terms and abbreviations used throughout the thesis.

Air Traffic Control = A system used to monitor and direct air traffic.

Alternate Hypothesis (H_1) = *as related to a statistical test* - "The hypothesis that is adopted when H_0 [Null Hypothesis] is rejected. Usually the same as the research hypothesis [i.e., the hypothesis the experiment was designed to investigate]" (Howell, 1985, p. 149).

ASRS *see Aviation Safety Reporting System*

ATC *see Air Traffic Control*

Aviation Safety Reporting System = A voluntary, confidential incident reporting system managed by NASA for the FAA.

CDU *see Control and Display Unit*

Control and Display Unit (CDU) = A device consisting of a display screen and a keyboard which allows the pilot to interact with the FMS.

Chi Square (χ^2) = A statistical test "used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories" (Sharp, p. 181).

Critical Period = All the events that take place between the time that the "desired state" was defined and the time that the flightcrew became aware that the desired state was not or would not be met given the current trend of activities.

Critical Value = *as related to a statistical test* - "the value of a test statistic at or beyond which we will reject H_0" (Howell, 1989, p. 150).

Descriptive Analysis = An analysis of data in which the set of data in the sample is described. Often descriptive statistical methods such as mean, medium, and range are used in this type of analysis.

Effect Size = *as related to a statistical test* - "the degree to which the phenomenon is present in the population" or "the degree to which the null hypothesis is false" (Cohen, pp. 9-10).

Expected Value = *as related to a statistical test* - The number of occurrences that would be expected if the null hypothesis was true. The expected value is derived from "the average value calculated for a statistic over an infinite number of samples" (Howell, p. 62).

FAA *see Federal Aviation Administration*

Federal Aviation Administration = The body of the U.S. government with primary responsibility for safety in civil aviation.

Flight Management System (FMS) = An advanced technology aircraft system used to perform a number of functions such as: determination of position, velocity and wind, trajectory determination, and computation of map and situation data for display. The FMS communicates with other avionics systems on the aircraft as well as with the pilot via the Control and Display Unit (CDU).

FMS *see Flight Management System*

Homogeneity = *as related to statistical samples* - All the parameters of the samples to be compared are the same except for the variables in which the analyst is interested.

Inertial Navigation System (INS) = A system on the aircraft that keeps track of its movement in the three spatial axes.

Inferential Analysis = An analysis of data in which methods are used to draw conclusions about the whole population based on a sample of that population. A common inferential statistical method used for such an analysis is the Chi Square (χ^2) test.

INS *see Inertial Navigation System*

Large Transport = Commercial aircraft are differentiated by arbitrary gross takeoff weight ranges. Large transport aircraft have a gross takeoff weight range of 150,001 to 300,000 lbs.

Medium-Large Transport = Commercial aircraft are differentiated by arbitrary gross takeoff weight ranges. Medium-large transport aircraft have a gross takeoff weight range of 60,001 to 150,000 lbs.

National Transportation and Safety Board (NTSB) = The agency responsible for investigating civil aviation accidents occurring in the U.S. and for providing U.S. Accredited Representatives to non-U.S. accident investigating boards when necessary. The NTSB also is responsible for issuing safety recommendations to the FAA aimed at preventing future accidents.

NTSB *see National Transportation and Safety Board*

Nominal Data = Data in which numbers are used to distinguish among objects. For example, an object is classified as a '1' if the phenomenon exists and classified as a '2' if the phenomenon does not exist. The numbers are not used for ordering the objects and the intervals between the numbers are not meaningful.

Null Hypothesis (H_0) = *as related to a statistical test* - "The statistical hypothesis tested by the statistical procedure. Usually a hypothesis of no difference or of no relationship" (Howell, 1985, p. 144).

Power = *as related to a statistical test* - "The probability that the test will yield statistically significant results" (Cohen, 1988, p. 1). The power is determined by the statistical test used, the sample size, and the effect size to be detected. A standard convention is to set the power to 0.80, which denotes an 80% chance that the test will yield statistically significant results.

Significance Level = *as related to a statistical test* - "The standard of proof that the phenomenon exists, or the risk of mistakenly rejecting the null hypothesis" (Cohen, 1988, p. 4).

Task Management (TM) = The function in which the human operator manages his/her available sensory and mental resources in a dynamic, complex, safety-critical environment in order to accomplish the multiple tasks competing for a limited quantity of attention.

Task Prioritization Error = A type of TM error in which the human operator misprioritizes the tasks to be performed.

TM *see Task Management*

Type I error = *as related to a statistical test* - Occurs when the null hypothesis is rejected when it is true.

Type II error = *as related to a statistical test* - Occurs when the null hypothesis is not rejected when it is false.

Widebody Transport = Commercial aircraft are differentiated by arbitrary gross takeoff weight ranges. Widebody transport aircraft have a gross takeoff weight range of anything over 300,000 lbs.

Appendix B: ASRS Incident Report Samples

Each sub-sample is composed of 70 ASRS Incident Reports randomly chosen without replacement from the total reports submitted in the appropriate time period that met the population parameters described in Table 4.1. The reports are listed by their ASRS accession number.

Sub-Sample of Traditional Technology Reports: 1988-1989

There were a total of 528 reports from traditional technology aircraft submitted in 1988 and 1989.

81222	88193	100086	113719	125807
81617	88219	100796	115990	125845
81801	88224	104037	116244	127316
81940	88307	104744	117048	127493
82048	90110	104858	117167	127900
82059	90360	105173	117198	128634
82491	90628	105248	117395	128853
82770	90644	105667	118178	129276
84116	92435	106358	120009	129411
84790	95887	106439	120411	130372
84894	98616	107565	120428	130744
84984	98883	109829	122089	130893
86711	99498	109856	123326	131532
86967	99813	113638	125493	132100

Sub-Sample of Traditional Technology Reports: 1990-1991

There were a total of 628 reports from traditional technology aircraft submitted in 1990 and 1991.

136649	147250	165703	175627	188220
137401	148684	166588	175659	189142
137721	148927	167336	175733	189690
138162	153477	167676	175772	190180
138655	156185	167813	177042	190616
139723	157038	169639	178094	190989
140177	159911	169924	179729	191039
140498	160314	170300	179807	192035
141490	161024	171136	181971	192839
142316	161818	171884	182353	194152
144473	163352	173051	182436	196447
145239	163791	173340	184987	196627
145245	164350	174467	186506	197365
146466	164501	175211	186532	197432

Sub-Sample of Traditional Technology Reports: 1992-1993

There were a total of 478 reports from traditional technology aircraft submitted in 1992 and 1993.

198353	207203	219487	237390	246252
198388	207284	220650	237947	247076
198938	208080	222178	239629	247600
199965	209498	222905	240589	247639
200803	209868	224640	240792	248001
201445	210234	226934	240949	248448
202238	210671	229062	241812	248728
202491	210678	229455	243141	249864
202728	211410	231964	243275	250768
203269	213286	232676	243853	251453
203357	216038	233685	244247	253220
203935	217103	236509	244448	253725
206547	217283	237221	244544	253763
207085	219383	237241	245041	256183

Sub-Sample of Advanced Technology Reports: 1988-1989

There were a total of 489 reports from advanced technology aircraft submitted in 1988 and 1989.

80482	90859	101423	111759	124169
81303	91425	101462	112142	124641
81559	91811	101607	112293	125395
81969	92106	106051	112745	125486
82202	92507	106566	113574	125761
83555	93803	108885	115112	126173
84781	95089	109432	115619	127203
85157	96184	110502	117145	128009
85170	96533	110569	118434	129253
86887	96787	110586	119740	129470
88353	99260	110593	121576	129501
88706	99620	110898	121749	129622
88976	100618	111196	124072	130700
89875	101335	111490	124108	130973

Sub-Sample of Advanced Technology Reports: 1990-1991

There were a total of 789 reports from advanced technology aircraft submitted in 1990 and 1991.

133702	146026	169279	179717	188019
134267	146918	169519	180669	188205
134903	147561	169533	181421	188583
135948	148721	169557	181530	189426
136799	149918	170238	182391	189664
137454	152370	170282	183653	190642
138178	153748	170508	183696	191157
140270	156758	170820	185071	191573
142369	162087	172204	185165	191629
143474	162485	172451	185585	193696
143616	164998	172596	186388	195077
143876	165672	177458	187052	196409
144367	167492	179046	187405	197311
145076	168743	179661	187921	197363

Sub-Sample of Advanced Technology Reports: 1992-1993

There were a total of 837 reports from advanced technology aircraft submitted in 1992 and 1993.

198895	214637	224908	234699	245327
200015	216140	224965	235856	247484
201029	216617	225315	236067	250641
201714	217197	225730	237132	251083
202757	218490	227317	238246	251602
207252	218531	227644	238399	251949
208115	218667	227692	238611	252685
210140	218771	227825	239324	252974
210587	219832	230962	242363	253435
210807	221121	231461	242789	254092
211961	221890	233213	244579	254105
213659	222283	233680	244978	254544
214015	223672	233905	244994	258555
214182	224291	234324	245283	258563