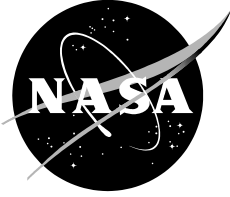# Checklists and Monitoring in the Cockpit: Why Crucial Defenses Sometimes Fail

R. Key Dismukes
NASA Ames Research Center, Moffett Field, CA

Ben Berman
San Jose State University Foundation, San Jose, CA

July 2010

Since it's founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

• TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

• TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

• CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

• CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

• SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

• TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

• Access the NASA STI program home page at http://www.sti.nasa.gov

• E-mail your question via the Internet to help@sti.nasa.gov

• Fax your question to the NASA STI Help Desk at (301) 621-0134

• Phone the NASA STI Help Desk at (301) 621-0390

• Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076-1320

Checklists and Monitoring in the Cockpit: Why Crucial Defenses Sometimes Fail

R. Key Dismukes
NASA Ames Research Center, Moffett Field, CA

Ben Berman
San Jose State University Foundation, San Jose, CA

Acknowledgements

Available from:

NASA Center for AeroSpace Information
7121 Standard Drive
Hanover, MD 21076-1320
(301) 621-0390

NASA Center for Aerospace Information
7121 Standard Drive
Hanover, MD 21076-1320
(301) 621-0390

Table of Contents

Acronyms

APU .... auxiliary power unit

ASAP .... Aviation Safety Action Program

ATC .... air traffic control

CDU.................... control display input

CRM .... crew resource management

FAA .... Federal Aviation Administration

FL .... flight level

FMS .... flight management system

FOM.................... flight operations manual

FOQA .... Flight Operations Quality Assurance

IMC..................... instrument meteorological conditions

IOE...................... initial operating experience

LOSA.................. Line Operations Safety Audits

MCP.................... mode control panel

NTSB .... National Transportation Safety Board

SOP..................... standard operating procedures

TCAS .... traffic collision avoidance system

TEM.................... threat and error management

Checklists and Monitoring in the Cockpit:
Why Crucial Defenses Sometimes Fail

R. Key Dismukes and Ben Berman

*Checklists and monitoring are two essential defenses against equipment failures
and pilot errors. Problems with checklist use and pilots' failures to monitor
adequately have a long history in aviation accidents. This study was conducted to
explore why checklists and monitoring sometimes fail to catch errors and
equipment malfunctions as intended. Flight crew procedures were observed from
the cockpit jumpseat during normal airline operations in order to: 1) collect data
on monitoring and checklist use in cockpit operations in typical flight conditions;
2) provide a plausible cognitive account of why deviations from formal checklist
and monitoring procedures sometimes occur; 3) lay a foundation for identifying
ways to reduce vulnerability to inadvertent checklist and monitoring errors; 4)
compare checklist and monitoring execution in normal flights with performance
issues uncovered in accident investigations; and 5) suggest ways to improve the
effectiveness of checklists and monitoring. Cognitive explanations for deviations
from prescribed procedures are provided, along with suggestions for
countermeasures for vulnerability to error.*

1. Executive Summary

Checklists and monitoring are two essential defenses against equipment failures and pilot errors.
Problems with checklist use and pilots' failures to monitor adequately have a long history in aviation
accidents.

A typical airline flight requires a great number of routine flight control inputs and switch actions and
frequent reading and verification of visual displays. Many of these actions are governed by formal
procedures specifying the sequence and manner of execution, after which checklists are used to
bolster reliability. Throughout the flight, pilots are required to monitor many functions, the state of
aircraft systems, aircraft configuration, flight path, and the actions of the other pilot in the cockpit.
Thus, the number of opportunities for error is enormous, especially on challenging flights, and many
of those opportunities are associated with checklists and monitoring—themselves safeguards
designed to protect against error.

Our study was conducted to explore why checklists and monitoring sometimes fail to catch errors
and equipment malfunctions as intended. In particular, we wanted to: 1) collect data on monitoring
and checklist use in cockpit operations in typical flight conditions; 2) provide a plausible cognitive
account of why deviations from formal checklist and monitoring procedures sometimes occur; 3) lay
a foundation for identifying ways to reduce vulnerability to inadvertent checklist and monitoring
errors; 4) compare checklist and monitoring execution in normal flights with performance issues
uncovered in accident investigations; and 5) suggest ways to improve the effectiveness of checklists
and monitoring.

1.1 Study Approach

Our approach was to observe flight crew procedures from the cockpit jumpseat during normal airline operations involving diverse aircraft types. Although we focused primarily on deviations from the idealized prescription for checklist execution and monitoring found in Flight Operations Manuals (FOMs), we attempt to put these deviations in context with examples of effective, often exemplary performance—which is far more common.

The second author (Berman) observed 60 normal operational flights from the cockpit jumpseat at three airlines (Table 1). One airline was a major U.S. flag carrier, one was a major U.S. domestic carrier[1], and one was a major foreign flag carrier. We attempted to record every observable deviation, even the most minor, including deviations that may have been necessitated by operational conditions. Our objective was to provide as complete an account as possible of the full range of deviations that occur under normal operating conditions so that (1) reasons for deviation can be determined, and (2) deviations that are problematic can be identified and addressed. As much as possible we avoid the value-laden term "error" in this report because, at least in some cases, deviation may have been appropriate, and in other cases may have been difficult to avoid.

1.2 Results and Discussion

Eight hundred ninety-nine deviations were observed (194 in checklist use, 391 in monitoring, and 314 in primary procedures). Deviations in the three major categories were sorted into types of deviation within the category (Tables 5,6, and 7) for further analysis. Somewhat speculative, but arguably plausible, cognitive accounts were developed for vulnerability to each category of deviation, based on analysis of the tasks being performed, the nature of cognitive skills, situational factors, and organizational factors.

Table 2 shows the number of deviations crews made per flight (means: checklists, 3.2; monitoring, 6.5; primary procedures, 5.2; total, 15.0). Variability across flights was quite large; for example, no primary procedure deviations were detected on one flight but 21 were observed on another flight (see Figure 1 on page 11). The distribution of the number of deviations per flight was substantially skewed to the right (a long tail of higher deviation rates) for all deviation categories. For example, on 31 flights 0–2 checklist deviations were observed, but on the other 29 flights 3–13 were observed. Thus a subset of flights produced a disproportionate number of deviations.

The number of deviations per flight should be considered in the context of the number of opportunities for deviation. For example, one airline used 10 checklists with a total of 197 challenge items plus response items. Several types of deviation could be made for each item (failure to respond, using non-standard phraseology, failure to look at item checked, etc). Thus, even if we considered all of these deviations to be errors, the rate of occurrence in terms of errors per opportunity was probably well under one percent, which is in the ballpark for many forms of skilled human performance. Put another way, in the vast majority of cases, checklists and monitoring were performed appropriately.

Rather than creating a deviation taxonomy *a priori*, or using one of the several error taxonomies that have been proposed for cockpit operations, we sorted each of the three deviation categories (checklist, monitoring, and primary procedure) into types according to similarity in operational

---

[1] Only two flights were observed at this airline because of scheduling and logistics difficulties.

aspects. Checklist deviations clustered into six types: flow-check performed as read-do; responding without looking; checklist item omitted, performed incorrectly, or performed incompletely; poor timing of checklist initiation; checklist performed from memory; and failure to initiate checklists (in order of number of occurrences; Table 5). The first two types accounted for nearly half of the checklist deviations observed.

Monitoring deviations grouped in three clusters: late or omitted callouts, omitted verification, and not monitoring aircraft state or position (Table 6). Over half of the monitoring deviations were late/omitted callouts, most of which (140) were the "1,000 feet to go" call, required as the aircraft approaches level-out altitude. Much more serious were omitted callouts during 11 approaches that were unstabilized, eight of which remained unstabilized beyond the final gate.

Although this study focused mainly on checklist use and monitoring deviations, additional data on primary procedure deviations provide context and allowed us to examine how effective checklists and monitoring were at trapping primary procedure errors. We grouped the 15 types of primary procedure deviations into six areas: 1) coordination within the crew or with ATC; 2) use of automation; 3) approach stabilization; 4) path and airspeed control; 5) configuration of systems or flight controls; and 6) planning and execution (Table 7). By far the most common deviations were failure to properly configure systems (62 instances), poor planning for contingencies (57 instances), poor coordination between the pilots (56 instances), and problematic use of the FMS (40 instances). Most of these deviations appeared to be inadvertent and can properly be described as errors.

We discuss at considerable length the cognitive, operational, and organizational factors that probably contributed to each type of deviation from SOP within the three categories. We also analyzed the data for possible influence of factors reported in previous studies to be associated with crew error. In contrast with an NTSB study of accidents attributed to crew error, we did not find that flights running late produced more deviations. However, consistent with previous studies, we did find that crews on their first flight together or on their first day of flying together made substantially more deviations. First officers and captains in their first year in aircraft type and seat position did not make more deviations than pilots with more than one years in type and position, however the three airlines at which we observed operations hire only pilots with substantial experience; thus this result might not apply to smaller airlines that hire pilots with substantially less experience.

Only 18% of deviations—even those that were clearly errors—were trapped (caught and corrected) or even discussed, a disquieting finding. In comparison, Klinect et al. (1999) reported that 36% of errors observed in LOSA were trapped, and Thomas and Petrilli (2006) reported 63% were detected and actively managed in a flight simulation study. Our lower trapping rates probably reflect multiple factors, one of which is that we observed actual line operations, in which operational pressures and opportunities for error are not fully captured by simulations. Also, the lower trapping rate we observed may reflect the fact that we deliberately recorded even very minor deviations, which is probably not true of most LOSAs. The percent of deviations trapped varied greatly across deviation types. In general, primary procedure deviations were more often caught: 35% versus 14% of checklist deviations and 6% of monitoring deviations. It is not surprising that monitoring deviations were least likely to be caught, since monitoring can be considered a final defense against primary errors (Sumwalt et al, 2002). Very large differences in trapping occurred among the types of deviation within each category. Only one of 113 verification omissions, 12 of 211 late or omitted callouts, and one of 48 flow-checks performed as read-do were trapped. In contrast, 25 of 33 failures

3

of crew-ATC coordination, 14 of 18 MCP deviations, and 32 of 62 system configuration deviations were trapped.

These large differences in trapping of different deviation types may reflect how conspicuous the consequences of the deviation are to the pilots and other personnel. Also, whether one pilot challenges a deviation by the other pilot may reflect how dangerous the deviation is perceived to be. In some situations, even when one pilot detects the other's deviation, it may be difficult or awkward to challenge the deviation. For example, "one thousand to go" calls must be made shortly before the altitude alerter chimes, and it is not clear to the flying pilot until the chime sounds whether the monitoring pilot will make the call. (At some airlines, the flying pilot makes this callout.) Further, the monitoring pilot—especially if a first officer—must consider whether frequently pointing out deviations that are unlikely to be consequential will create a tense cockpit. Similarly, a captain must be selective about challenging errors made by the first officer in order to avoid micromanaging the flight deck, which undercuts open communication.[2] On the other hand, in some situations it is difficult for a pilot to assess in real time whether an error will have significant consequences. Any missed callout or verification removes the power of that action to trap errors and prevent undesired aircraft states.

Captains in the monitoring pilot role were more than twice as likely to trap deviations made by the flying pilot than first officers in the monitoring pilot role (27.9% versus 12.1%), which points to the need to develop ways to encourage first officers to challenge when appropriate.

Based on a sample of slightly more than half of the flights that we evaluated as to consequences, eighty-nine percent of the observed deviations had no discernable outcome other than an arguably small reduction in the efficacy of safeguards. For example, even though pilots sometimes failed to make the "thousand feet to go" call the autopilot leveled the aircraft at the correct altitude, though of course if the FMS or MCP had been set up incorrectly, the aircraft might not have leveled off. The fact that the great majority of deviations do not lead to serious consequences suggests that the overall system of multiple, overlapping safeguards works fairly well. However, nine percent of deviations led to an undesired aircraft state, and two percent led to subsequent deviations.

We observed 45 instances of undesired aircraft state of diverse sorts: deviations in airspeed, heading, or vertical path; incorrect heading set for takeoff; incorrect configuration of controls or systems; flight attendants not seated when required by SOP; unstabilized approaches and landing from unstabilized approaches; inadequate terrain separation, etc. (Table 12). Clearly these undesired states—some resulting from multiple deviations--were more serious than the outcome of most deviations in that the potential for an accident was greater.

1.3 Countermeasures

We developed a set of countermeasures that we believe would substantially reduce pilots' vulnerability to deviating from SOP:

1.3.1 Cockpit Procedures and Organization Policies

Suggestion: Formalize monitoring and challenging requirements and procedures.
Suggestion: Minimize checklist items involving multiple components and specify responses for each component.

---

[2] We are indebted to a senior airline captain for pointing this out.

Suggestion: Evaluate error vulnerability of existing procedures and strengthen them.
Suggestion: Organizations should periodically review cockpit operating procedures to identify and relieve "hotspots" in which prospective memory and concurrent task demands are high and interruptions are frequent.
Suggestion: Organizations should systematically analyze the entire body of explicit and implicit messages given their pilot corps to balance competing goals.
Suggestion: Organizations should examine the role of organizational procedures in vulnerability to error in the cockpit (as well as errors in the cabin, dispatch center, and maintenance hangar).

1.3.2 Training, Checking, and Mentoring

Suggestion: Pilots should be trained on their inherent vulnerability to checklist and monitoring errors, and on procedural measures and practical techniques to counter it.
Suggestion: Reinforce the responsibility of monitoring pilots to challenge deviations.
Suggestion: Develop techniques to provide detailed feedback to pilots on checklist and monitoring performance.
Suggestion: Place greater emphasis on checklist use and monitoring in air carrier flight standards (line checking) programs.
Suggestion: Develop formal mentoring programs for new first officers.

1.3.3 System Design

Existing systems, such as mechanical and integrated electronic checklists, already used in some aircraft, can reduce vulnerability to some of the checklist deviations observed in this study. The next generation of integrated electronic checklists, with expanded ability to sense the status of flow/checklist items, will further protection, and artificial intelligence may provide intelligent agents to help pilots catch deviations. However, although cockpit automation comes with many benefits, it can also introduce new problems (Billings, 1997; Sarter and Woods, 1994), such as automation mode confusion and automation complacency.

Suggestion: Research is needed to develop ways to help pilots stay in the loop on system status, aircraft configuration, flight path, and energy state. These new designs must be intuitive and elicit attention as needed, but minimize effortful processing that competes with the many other attentional demands of managing the flight.

1.4 Conclusion

Although this study focused on deviations from prescribed procedures, these deviations must be understood in context. The vast majority of the actions of the observed crews were correct and effective and demonstrated required skills. Given the large numbers of opportunities for deviation, the deviation rates were probably well below one percent. We observed many examples of exemplary performance and of effective techniques used to manage the challenges of cockpit operations.

Even though modern airlines operate at extremely high levels of safety, the very fact that the level of safety is so high makes it difficult to detect when safety begins to erode. The tendency of any highly organized system is to become less well organized (using a metaphor from physics, entropy increases); thus, constant effort is required to maintain safety. The industry is under extreme pressure to cut costs, and the consequences of changes to training and procedures do not always show up immediately.

Our findings point to things that can be improved. In particular, trapping of errors and other deviations appears not to be operating at the level generally assumed. Most people in the airline industry now recognize that it is impossible to eliminate all human error, and that it is necessary to help pilots detect and manage errors before they become consequential. Threat and error management (TEM) programs are now fairly common, and many airlines address the need for cockpit monitoring. Yet these well-intentioned efforts appear to be falling short. The countermeasures we suggest could provide a path to improvement.

2. INTRODUCTION

On 14 August 2005, a Boeing 737 operated by Helios Airways departed Larnaca, Cyprus headed for Prague. As the aircraft climbed through 16,000 feet, the captain radioed the company operations center and reported a take-off configuration warning and an equipment cooling-system problem. Passenger oxygen masks automatically deployed at 18,200 feet, and communication between the flight crew and ground facilities ended when the aircraft passed through 28,900 feet and then leveled out at flight level (FL) 340 on autopilot. (FL 340 is approximately 34,000 feet above sea level.) The 737 was intercepted by two F-16s from the Hellenic Air Force, whose pilots attempted visual contact with the flight crew. One of the 737 pilots appeared unconscious and the other was not visible. After cruising on its pre-programmed course for three hours, the 737's engines flamed out and the aircraft crashed, killing all 121 persons aboard (AAIASB, 2006).

The subsequent investigation determined that the 737's pressurization system had been set to the manual position (apparently by maintenance personnel) and had not been re-set to the automatic position, as required by the airline's formal procedures, by the flight crew. The pilots did not detect the mis-setting when performing their preflight procedures and did not catch the oversight when running the Before Start checklist and the After Takeoff checklist. Apparently the pilots then mistook the cabin altitude warning for a takeoff configuration warning, became preoccupied with this erroneous interpretation as well as an equipment cooling system warning (associated with the depressurization), and allowed the aircraft to continue climbing until they passed out from lack of oxygen.

This accident was not unique. Problems with checklist use and failures to monitor aircraft systems adequately have a long history in aviation accidents (Turner & Huntley, 1991; Turner, 2001; NTSB, 1994). Degani and Wiener (1993) published a qualitative study that identified forms of error in use of normal checklists[3] and discussed issues of design and use. Problematic performance included bunching several checklist items in single challenges and responses, performing flow-then-check items as read-do, failing to call checklists complete, erroneously perceiving a mis-set item as correctly set, failing to cross-check items set by one pilot, and failing to complete items or entire checklists (the latter often due to interruptions and distractions). Degani and Wiener (1993, 1994) analyzed problems with the design of many normal checklists and provided human factors guidelines for improving design.

---

[3] *Normal* checklists are used in routine flight operations to ensure that controls and systems are correctly set and are operating properly, in contrast to *non-normal*, or *emergency* checklists, which are used to help pilots identify malfunctions and respond appropriately.

The Degani and Wiener study caught the attention of the airline industry (Gross, 1995), and many airlines modified their checklists using the study's guidance to address concerns about checklist design and execution. However, little research has been published examining pilots' performance of checklist procedures in recent years, even though the airline industry has seen substantial change in the past two decades. These changes, the Helios accident, a SpanAir accident in 2008—in which an MD-82 crashed when the flight crew attempted to take off with flaps not set—and numerous ASRS reports of takeoffs rejected because of configuration warnings that critical controls were not set properly suggest that an observational study of how checklists are currently being used in routine line operations should be conducted.

Robert Sumwalt and colleagues called the attention of the aviation industry to the importance of monitoring as a defense against threats and errors (Sumwalt, 1999; Sumwalt, Thomas, & Dismukes, 2002, 2003). Monitoring refers to the responsibility of pilots to keep track of the aircraft's position, course, and configuration; the status of the aircraft's systems[4]; and the actions of the other pilots in the cockpit. Often, monitoring must be performed concurrently with other tasks such as operating aircraft controls, making data entries, and communicating, and this, unfortunately, may lead pilots to think of monitoring as a secondary task. The reality is that lapses in monitoring have played a role in many aviation accidents. A National Transportation Safety Board (NTSB) study found that inadequate monitoring/challenging played a role in 84% of major airline accidents attributed to crew error over a 12-year period (NTSB, 1994). (The accident reports did not provide the kind of information that would have been required to distinguish the relative contributions of monitoring lapses and *challenging* lapses—the latter being the failure of a pilot to call an observed error to the attention of the pilot making the error.). Most of these lapses were secondary failures to catch primary errors that the NTSB considered to be causes of the accidents. Similarly, the Flight Safety Foundation found that 63% of approach and landing accidents involved inadequate monitoring and cross-checking (FSF, 2010), and the International Civil Aviation Organization found inadequate monitoring to be a factor in 50% of controlled flight into terrain accidents, (ICAO, 1994).

In 2003, the Federal Aviation Administration (FAA) expanded its advisory circular on standard operating procedures to provide guidance on monitoring procedures (FAA, 2003). Consistent with this guidance, in recent years many airlines have changed the title of the pilot not flying to *monitoring pilot* and have revised flight operations manuals to explicitly describe at least some monitoring duties.

Thus, both monitoring and checklists are well established as crucial defenses against threats and errors. Yet, as the Helios and SpanAir accidents illustrate, these defenses sometimes still fail. A previous review of airline accidents attributed to crew error revealed that weakness in checklist use and monitoring, sometimes leading to fatal outcomes, are not isolated problems (Dismukes, Berman, & Loukopoulos, 2007). Also, a detailed flight simulation study of experienced pilots' monitoring of automation mode annunciations found that failures to detect mode changes were common (Sarter, Mumaw & Wickens, 2007). However, to our knowledge, no direct observational study of monitoring and checklist performance in actual flight operations has been published since Degani and Wiener (1993).

---

[4] The most recent generation of airliners uses centralized alerting systems that relieve pilots of much of the need to directly monitor each system; however, pilots must still periodically scan the integrated display of this centralized system and be aware of systems issues that may not be monitored automatically.

The airline industry has changed substantially in several ways in the last decade or so. Economic pressures have become quite severe; some airlines have gone out of business or been acquired by other airlines, and all airlines have had to institute severe cost-cutting measures to survive. Security measures in the wake of 9/11 have changed some aspects of flight operations. In some segments of the airline industry pilots are being hired with far less experience than in recent decades, and most of these pilots lack the military flying experience that was previously common in the U.S. industry. Cockpits are increasingly automated. All of these changes have the potential to affect how pilots are trained and how they execute cockpit procedures.

A typical airline flight requires a great number of routine flight control inputs and switch actions and frequent reading and verification of visual displays. Many of these actions are governed by formal procedures specifying the sequence and manner of execution, after which checklists are used to bolster reliability. Throughout the flight, pilots are required to monitor many functions, the state of aircraft systems, aircraft configuration, flight path, and the actions of the other pilot in the cockpit. Thus, the number of opportunities for error is enormous—especially on challenging flights, and many of those opportunities are associated with two safeguards themselves designed to guard against error: checklists and monitoring. The impressive safety record of airline operations in developed countries is testament that pilots perform the vast bulk of procedures correctly, neutralizing threats and averting potential consequences of errors. However, maintaining the safety of any highly ordered system—an aircraft or the entire air transport system—is a bit like balancing on a ball; constant effort is required to counter the many forces that would disorder the system.

With this context, our study was conducted to explore why checklists and monitoring sometimes fail to catch errors and equipment malfunctions as intended. In particular, we wanted to: 1) collect data on monitoring and checklist use in cockpit operations in typical flight conditions; 2) provide a plausible cognitive account of why deviations from formal checklist and monitoring procedures sometimes occur; 3) lay a foundation for identifying ways to reduce vulnerability to inadvertent checklist and monitoring errors; 4) compare checklist and monitoring execution in normal flights with performance issues uncovered in accident investigations; and 5) suggest ways to improve the effectiveness of checklists and monitoring. Our approach was to observe flight crew procedures from the cockpit jumpseat during normal airline operations involving diverse aircraft types. Although we focused primarily on deviations from the idealized prescription for checklist execution and monitoring found in Flight Operations Manuals, we attempt to put these deviations in context with examples of effective, often exemplary performance—which is far more common.

3. METHOD

The second author (Berman) observed 60 normal operational flights from the cockpit jumpseat at three airlines (Table 1). One airline was a major U.S. flag carrier, one was a major U.S. domestic carrier[5], and one was a major foreign flag carrier. Since 11 September 2001, researchers' access to airline cockpits during flight has been severely restricted by security precautions. However, the second author is an airline pilot with considerable experience as an observer for Line Operations Safety Audits (LOSA) and was able to get permission to fly in the jumpseat.

---

[5] Only two flights were observed at this airline because of scheduling and logistics difficulties.

Thirty-nine of the 60 observation flights were conducted as part of a LOSA. The remaining 21 observations were done in a manner similar to LOSA; crews received letters jointly signed by company management and the union explaining the purpose of the study and encouraging cooperation. Two crews were observed twice, and one other pilot was observed paired with two different pilots. We had planned to observe a larger number of crews twice to compare performance variability within and between crews; however, difficulties in scheduling jumpseat observations and the practice of switching pilot flying and pilot monitoring roles between legs made this plan impractical.

Observations were made in six types of aircraft (Table 1). Before flight the observer studied the airline's flight operations manual (FOM) for the aircraft type, which describes cockpit procedures in detail, including the checklists and monitoring procedures. The observer introduced himself to the flight crew either while they were waiting to board the aircraft or after they were seated in the cockpit and asked permission to observe the flight. All crews gave permission to be observed. The observer attempted to be as unobtrusive as possible during the flight; however, because jumpseat occupants are technically members of the flight crew, he was obliged to raise any concerns if significant flight safety issues arose, which happened only a few times.

During the flight the observer took free-form notes on a small notepad. A printed observation guide was prepared and reviewed before flight to help standardize observations; however, the guide was not consulted during flight to avoid distracting the flight crew by shuffling through sheets of paper. Observations were recorded about deviations from the company's procedures for checklist use, monitoring, and other procedures, as well as any circumstances that seemed relevant to the crews' execution of procedures (e.g., interruptions, high workload, or unusual circumstances). When a pilot deviated from a procedure prescribed in an FOM or aviation regulation, we tracked whether either crewmember identified the deviation, whether it was corrected, and whether it led to any consequences with potential to affect the outcome of the flight. We also tried to record successful and exemplary performance; however, pilots' deviations were easier to observe than things simply done right, which occurred much more frequently.

Because an important function of checklists and monitoring is to trap errors in aircraft operation (aircraft control, navigation, communication, and planning), we also recorded deviations in these "primary" operations. We made no distinction between inadvertent deviations (slips, lapses, and omissions) and intentional noncompliance because in most instances we could not infer intent with confidence.

During cruise the observer asked the crew whether they had previously been paired together on this trip or on any other flight, and recorded other information such as which pilot (captain or first officer) was the flying pilot and which was the monitoring pilot. Immediately after the flight the observer used his notes to write up a narrative in a standard format, and from these narratives specific occurrences were later coded, identifying (to the extent possible) each event, its antecedents, and its consequences.

We attempted to record every observable deviation, even the most minor, including deviations that may have been necessitated by operational conditions. Our objective was to provide as complete an account as possible of the full range of deviations that occur under normal operating conditions so that (1) reasons for deviation can be determined and (2) deviations that are problematic can be identified and addressed. As much as possible we avoid the value-laden term "error" in this report

because, at least in some cases, deviation may have been appropriate, and in other cases may have been difficult to avoid.

Deviations in the three major categories—checklist use, monitoring, and primary procedures—were sorted into types of deviation within the category (Tables 5, 6, and 7) for further analysis. Somewhat speculative, but arguably plausible, cognitive accounts were developed for vulnerability to each category of deviation, based on analysis of the tasks being performed, the nature of cognitive skills, situational factors, and organizational factors.

The reader will undoubtedly note the limitations of this observational method. Reliability cannot be assessed because only one observer could go on each flight. The observer has more personal experience in some of the procedures and types of aircraft used in this study than in others; thus we make no attempt to make comparisons among different airlines and among different aircraft types, and we must be cautious about interpreting the relative frequency of different types of deviations observed. Not all deviations are equally observable and some may not be observable at all. For example, the observer recorded whether pilots appeared to be looking at the items being checked on the basis of the direction the pilots' heads were turned, but this approach cannot detect instances in which heads were turned in the right direction but gaze[6] was not directed to the item, and, conversely, when heads were not turned exactly toward the item but gaze was directed eccentrically toward the item. Our goal was simply to obtain a substantial sample of deviations and relevant factors in a cross-section of routine airline flight operations, and this seems to have been achieved.

4. RESULTS

Eight hundred ninety-nine deviations were observed (194 in checklist use, 391 in monitoring, and 314 in primary procedures).

The captain was flying on 37 of the 60 flights and the first officer on the other 23. Because it is common practice for pilots to alternate the flying pilot and monitoring pilot roles, we did a chi-squared test and determined that the probability of this large a distribution imbalance occurring randomly was .07. Thus the larger percentage of flights operated by captains might have occurred through chance; however, captains may choose to fly more legs for reasons such as bad weather, and it may be that some captains in this study chose to fly when they learned the flight was going to be observed.

The mean duration of flight was 2.0 hours (standard deviation 1.6; median duration 1.3 hours). The number of deviations was not correlated with flight duration.

Table 2 shows the number of deviations crews made per flight (means: checklists, 3.2; monitoring, 6.5; primary procedures, 5.2; total, 15.0). Variability across flights was quite large; for example, no primary procedure deviations were detected on one flight but 21 were observed on another flight (Figure 1). The distribution of the number of deviations per flight was substantially skewed to the right (a long tail of higher deviation rates) for all deviation categories, and this was confirmed by computing skewness coefficients (checklists, 1.2; monitoring, 1.0, and primary procedures, 1.3). For

---

[6] Gaze is the direction the eyes are pointed. People generally turn their heads toward what they wish to see, but small adjustments are often made by moving the eyes eccentrically to the direction of the head.

example, on 31 flights 0-2 checklist deviations were observed, but on the other 29 flights 3-13 were observed. Thus a subset of flights produced a disproportionate number of deviations.
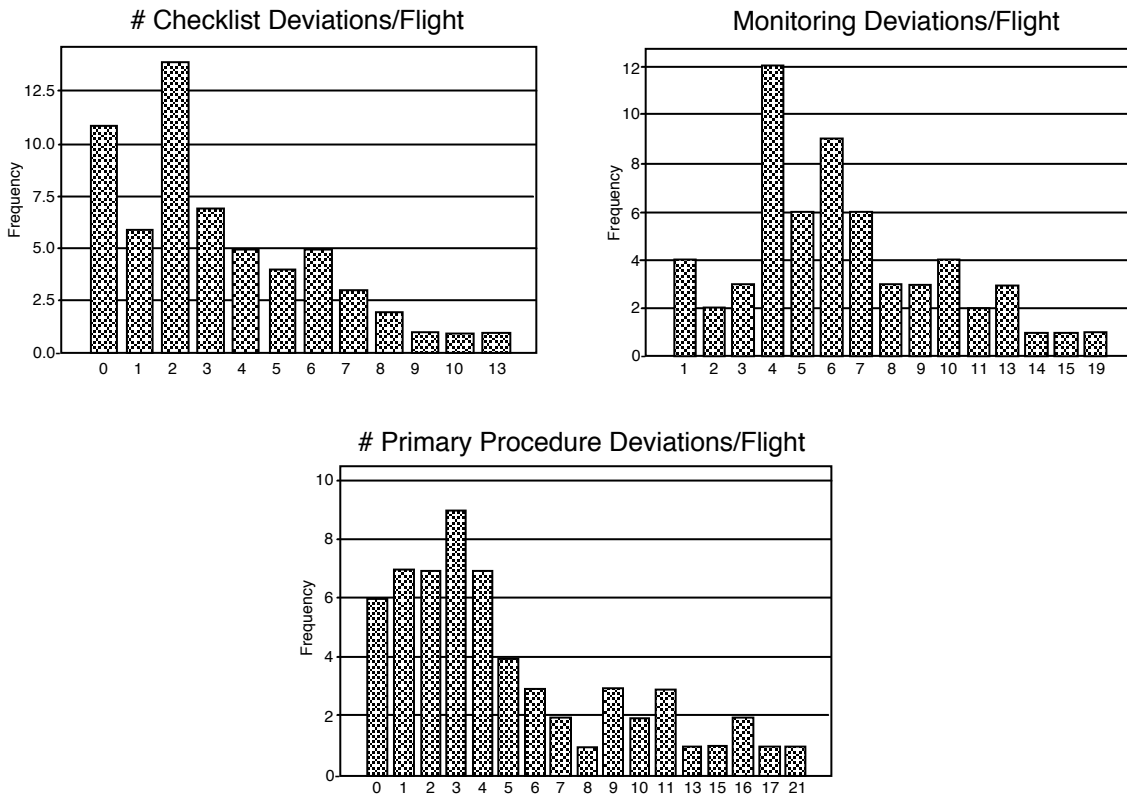
Figure 1. Distributions of deviations/flight.

The variability across flights in the number of deviations might result from several causes: differences among pilots in how they typically perform procedures, differences in conditions in which the flights were conducted, and differences in the observer's noticing deviations from one flight to the next, among other possibilities. Although we cannot clearly separate the relative roles of these factors, we did an analysis that suggests some of the variability lies in differences among pilots in how they typically perform procedures. The number of deviations in each category that crews made before takeoff was compared with the number of deviations in that category made during and after takeoff on the same flight on the assumption that how rigorous crews were in following procedures as written would be fairly consistent throughout the flight. We found that the number of deviations in each category before takeoff correlated with the number during and after takeoff: checklists (r = .54), primary procedures (r = .50), monitoring (r = .30), and total deviations (r = .54). All of these correlations were highly significant statistically, except for monitoring (p < .11)[7], suggesting that some of the variability in deviation rates among flights is due to differences in how each crew adhered to the prescribed manner of executing procedures.

---

[7] The standard criterion of p < .05 was used to assess statistical significance.

Table 3 shows the distribution of deviations across the phases of flight. Most deviations were made during the pre-taxi (19%), climb (23%), and descent (23%) phases. Checklist deviations were especially associated with pre-taxi, taxi-out, descent, and approach phases. Monitoring deviations were especially associated with climb and descent. Primary procedure deviations were more evenly distributed among phases but were most prominent in pre-taxi, cruise, and descent. The distribution of deviations may reflect in part the relative number of opportunities for deviation; for example, the number of checklist deviations is distributed roughly proportionally to the number of checklist challenge and response items in each phase of flight (Table 4). However, the number of deviations is clearly not simply a function of how long the flight phase lasts, because more errors in all three categories were made in relatively short phases than in cruise, which typically lasts longer than other phases.

Fifteen flights (25%) were operating late, and 42 (70%) were on time (data were not available on three flights). Crews were not significantly more likely to deviate on flights operating late.

4.1 Types of Deviations

Checklist deviations. For a typical air carrier checklist, the monitoring pilot reads ("challenges") each item from a printed card or electronic display, the flying pilot responds by checking that the item is correctly set and verbalizing a standard response, and the monitoring pilot cross-checks the item. For example, an item on the Landing checklist is "landing gear," to which the response is "down, three green." (Three green refers to the lights indicating that each gear is locked down.) While most checklists involve both pilots working cooperatively, the monitoring pilot performs all of the challenges and responses for some checklists. Also, while most checklists involve verbalization of challenges and responses, some are designed to be performed silently. Regardless of verbalization of the items of the checklist, under most standard operating procedures (SOP) there will be an explicit callout to initiate the checklist and also an explicit callout that the checklist has been completed.

Six types of checklist deviation were identified on the observed flights (Table 5). The most common type was performing a flow-check procedure as a read-do procedure (48 instances). At the observed airlines, most normal (as opposed to non-normal and emergency) checklist procedures require pilots to check and/or set items to the required position in a standard sequence (the "flow"), after which a checklist is run to ensure that the most critical items in the sequence have been performed correctly. For example, at one observed airline, during descent the flying pilot calls for the In-range checklist, which initiates a flow sequence by the monitoring pilot that includes, among other things, turning on the seat belt sign, checking the pressurization, and scanning the flight instruments for failure flags. After completing the flow, the monitoring pilot performs the checklist, which includes re-checking the first two items but not the instrument panel scan. If the monitoring pilot skips the flow and proceeds directly to the checklist, that scan and other items not specified on the checklist will be omitted. In contrast, with read-do checklists (used most commonly with emergency checklists, which are performed much less frequently) one pilot reads every item to be performed and after each item is read the pilot specified by the procedure sets or checks that item.

In 43 instances, a pilot either responded verbally to a challenge item without visually inspecting the item, responded verbally before inspecting the item or responded that the item was correctly set when in fact it was not. For example, a first officer did not look up from the checklist card to verify items on the overhead panel, and a captain responded "On" to the "APU Bleed" challenge when the

bleed was actually off. The latter incident may be an example of "looking without seeing," discussed later.

In 42 instances an item was omitted from a checklist, the response was incorrectly worded (e.g. "set" was stated rather than the numeric value required), one or more elements of a multi-item response were omitted or combined into a single response, or the checklist was not called "complete" after the last item was performed. In some instances the checklist item was deferred and later forgotten, in others the checklist was interrupted by some external agent or event and an item was overlooked, but in many cases an item was omitted when no external disruption of checklist execution was observed.

In 31 instances the flying pilot called for a checklist either at the wrong time or at a time that interfered with higher priority tasks, or the monitoring pilot self-initiated a checklist that had not been called for, pre-empting initiation at the proper time. For example, a first officer apparently forgot to call for the In-range checklist at 18,000 feet and only remembered to call for it at 10,000 feet, several minutes later. On another flight, a captain called for the Taxi checklist when the aircraft was approaching a runway intersection, causing the first officer to go head down at a time when it is crucial for both pilots to be looking outside the aircraft.

In 17 instances the monitoring pilot performed the checklist from memory instead of reading from the checklist card, and in 13 instances a pilot failed to call for a checklist to be initiated. (However, in 10 of these 13 instances the other pilot suggested running the checklist, so it was not omitted—see later sections of this report.)

Monitoring deviations. Three types of monitoring deviation were noted (Table 6). The most common (211 instances) was omitting a callout or making it late. By far the most common example of this subcategory (137 of the 211 instances) was omitting the "1000 feet to go" callout before altitude level-off, or making this call only after prompting by the automatic chime (in this latter case, we consider that a callout prompted by the chime did not provide an alert independent of the aircraft system, as designed by the standard procedures). A more serious example was omission of a callout required during unstabilized approaches; for instance, a monitoring pilot did not call out "Unstable" when the airspeed and thrust were not stabilized at 500 feet, a condition that mandates abandoning the approach and going around.

There were 113 instances of omission of a required verification. For example, descending through FL310 the flight received clearance to FL240, and while the first officer set and called out the new altitude, the captain was distracted by conversation and did not verify the new altitude on the primary flight display. In another instance, while climbing through transition altitude, the captain made the required callout and the pilots reset their altimeters to standard pressure, but neither pilot performed the required cross-check of the other pilot's altimeter.

Failure to monitor the aircraft state or position was noted 67 times. For example, a crew became occupied with planning weather avoidance and did not notice a fuel configuration EICAS message. In another instance the captain began his cruise cockpit panel scan early and did not monitor the autopilot's leveling of the aircraft at the assigned altitude.

Primary procedure deviations. The 314 instances of deviations in executing primary procedures were distributed among 15 types (Table 7). Of these, 103 instances involved coordination within the crew or with air traffic control (ATC), ground crew, or flight attendants (i.e., four types of

deviation); 66 involved systems or aircraft configuration (two types); 64 involved contingency or profile planning and execution (two types); 60 involved automation (three types); 11 involved deviations in path or airspeed (three types); and 10 involved unstabilized approaches. By far, the most common types were in the areas of:

1. Configuration of equipment and systems. For example, a captain turned on the engine anti-ice before the airplane entered the clouds in icing conditions, but he did not turn on the engine ignition. Just as the airplane entered the clouds the first officer noticed that the igniters were not turned on and he selected continuous ignition, thus trapping the captain's oversight.

2. Planning for, or responding to, contingencies. For example, near the end of one flight, at 6,500 feet, ATC transmitted, "Braking action fair reported by all types." The crew made no comments in response, and they did not calculate landing distance under the reported conditions. On another flight, neither pilot had the weather radar turned on while climbing in instrument meteorological conditions (IMC) and rain from 3,000 feet to FL200.

3. Crew-to-crew coordination. For example, at 15,000 feet a flight was cleared direct to a downline fix. The captain inputted and executed the route change without waiting for the first officer to confirm the change. Another flight was cleared to hold short of an intersecting runway, but neither pilot verbalized the hold-short restriction.

4. Data entry or use of the flight management system and mode control panel. For example, while a flight was climbing through 9,000 feet, the first officer accepted an ATC speed restriction of 270 knots above 10,000 feet. The captain programmed and executed a 270-knot climb; consequently, the airplane accelerated immediately to 270 knots, violating the regulation restricting speed to 250 knots below 10,000 feet. On another flight, the first officer did not arm the autopilot/flight director system to capture the ILS localizer as the flight neared the final approach course.

4.2 Crewmember Making the Deviation

Fifty-four percent of the total number of deviations (three categories combined) were made by captains and 46% by first officers (flying pilot and monitoring pilot roles combined) and deviations were evenly divided between flying pilot and monitoring pilot (captains and first officers combined). (Remember that captains were the flying pilots more often than first officers.) To compare the performance of captains with first officers in the flying pilot and monitoring pilot roles, it was necessary to examine only the deviations made in flight, because captains always taxi the aircraft on the ground, and to compute the number of deviations *per flight*. During the in-flight phases of the 60 observed flights, pilots made 604 deviations (74 checklist type, 331 monitoring type, and 199 primary procedure type). No significant differences in total number of *crew* deviations (captain deviations plus first officer deviations) occurred between flights in which the captain was the flying pilot and flights in which the first officer was the flying pilot (data not shown). Captains made slightly more deviations *per flight* (three categories combined) than first officers, both as flying pilots (4.6 vs. 4.2 deviations per flight) and monitoring pilots (5.5 vs. 4.4 deviations per flight); however, these differences were not statistically significant (Table 8). Also, no significant differences occurred between captains and first officers in the number of checklist, monitoring, and primary procedure deviations examined separately (data not shown).

Pilots who were flying together for the first time (eight out of the 56 flights for which data were available) made more total deviations than pilots who had previously flown together (22.4 total deviations per flight versus 14.2). This difference was marginally significant statistically.[8] Monitoring deviations, checklist deviations, and primary procedure deviations were all higher on first flights together, but only the primary procedure deviations were significantly different (10.1 deviations per flight versus 4.7; Table 9).

Pilots on their first day (though not necessarily their first flight leg) of flying together (18 of the 56 flights for which data were available) made more total errors than those not on their first day together (18.6 total deviations per flight versus 13.9). This difference was statistically significant. Monitoring, checklist, and primary procedure deviations were all higher on the first day together, but the difference was significant only for procedure deviations (7.8 deviations per flight versus 4.4).

Of the 60 observed flights, we obtained crewmember experience data for 57 flights. Five pilots were observed on two sequential flights, though not always paired with the same pilot as on the first flight; thus we had 109 observations of pilot performance to link to experience. We asked the pilots whether, at the time of observation, they were in their first year in the crew position (captain/first officer) and aircraft type (Airbus 320, etc.) being observed. Overall, 20 of the 109 observations (18 percent) were of pilots in their first year in position/type. Pilots in their first year in position/type made about the same number of deviations as those who were not. Following the NTSB analysis of 1994, we separately evaluated whether the first officer was in the first year in position and aircraft type. Of the 57 flights for which data were available, 17 (30 percent) were crewed by a first officer in his or her first year in that position and the observed aircraft type. We found no significant differences in the number of deviations made on these flights compared to those with more experienced first officers.

4.3 Outcomes of Deviations

Only 18% of deviations were corrected. We have no way of knowing which were not noticed and which were noticed but still not corrected. Of those that were corrected, most were caught by the other pilot (63%) and some were caught by the pilot making the deviation (17%) or other individuals (19%), such as air traffic controllers. Captains and first officers were equally likely to trap errors —Table 10. The number of errors trapped varied greatly with the category and type of deviation (Table 11). Twenty-two of 391 monitoring deviations (5.6%), 28 of 194 checklist deviations (14.4%), and 111 of 314 primary procedure deviation (35.4%) were trapped. These differences were statistically significant. Many of the types of deviation within the three categories had too few occurrences for statistical analysis of types to have appreciable power, but some differences stand out. Only one of 113 verification omissions, 12 of 211 late/omitted callouts, and one of 48 flow-checks performed as read-do were corrected. In contrast, 10 of 13 failures to initiate checklists, 25 of 33 failures of crew-ATC coordination, 14 of 18 mode control panel (MCP) errors, and 32 of 62 system configuration errors were caught. It may be that deviations easier to observe or those more likely to cause problems were more likely to be challenged.

---

[8] The difference was highly significant if equal variances are assumed ($p = .009$) but only marginally significant if equal variances are not assumed ($p = .087$). The latter assumption is more likely to be correct by Levene's test; $p = .054$.

12.1 percent of deviations made by the captain while acting as the flying pilot were pointed out by the first officer, whereas 27.9% of deviations made by the first officer as flying pilot were corrected by the captain, and this difference was statistically significant. These data do not reveal whether this large difference between pilots in the monitoring role was due to first officers' being less likely to notice or being less likely to challenge captains' deviations. In contrast to the monitoring pilot results, first officers while acting as the flying pilot responded to 8.5% of the captain's deviations as monitoring pilot, and captains in the flying pilot role responded to 6.2% of first officers' deviations as monitoring pilot. Thus the flying pilot was less likely to identify monitoring pilot deviations than vice versa, which is what one would expect, given the nature of the two roles.

After completing the original analysis of error trapping, we selected thirty-one flights for a detailed analysis of the outcomes of deviations that were not challenged. To provide a representative sample, these 31 flights comprised the first half of the flights observed during each of the three periods of field research (March 2005-February 2005; June-August 2007; and February-March 2008). (When the observations in one of these periods were not an even number, an extra observation was included in the sample.)

For the 31 sampled flights, 460 of the 518 deviations (88.8%) had no discernable outcome, other than reduction of efficacy of safeguards (e.g., when the crew failed to make the required callout of "1000 feet to go," the autopilot still correctly leveled the aircraft at the assigned altitude); 12 (2.3%) led to subsequent errors; and 44 (8.5%) resulted in an undesired aircraft state that required detection and correction by the flight crew to avoid a more negative outcome. (Undesired aircraft states are listed in Table 12.) The 44 deviations resulting in an undesired aircraft state were clearly errors; the most common of these errors were not monitoring aircraft state or position (5), failing to reject an unstabilized approach (4), systems misconfiguration (7), and inadequate contingency planning/execution (5)—see Table 13.

Unstabilized approaches were analyzed in more detail for the entire data set of 60 flights. Eleven unstabilized approaches occurred among the observed flights, including one or more at each of the three observed airlines. Two airlines used a standard stabilized approach criterion requiring rejecting the approach if not stabilized by and below 1000 feet. During the course of this study, the third airline established three gates for evaluating approach stability: an energy management gate (at 5 miles/1500 feet), a configuration gate (at 1000 feet) and a final gate at 500 feet. Deviations at the first two gates required correction while allowing the crew to continue the descent; deviation at the final gate required an "unstable" callout and a mandatory go around.

Twenty deviations were observed during the course of the 11 unstabilized approaches, involving both the flight path and aircraft configurations executed by the flying pilot and the callouts omitted by the monitoring pilot. Of these 11 approaches, in two flights the crew was able to stabilize the approach before the final gate (1,000 or 500 feet, as designated by the airline); in one flight the crew executed a go-around after the aircraft reached the final gate unstabilized; and in eight flights the crews continued to land even though unstabilized at the final gate. All eight flights were stabilized before landing, although the altitudes at which stabilized ranged from 200 feet to 900 feet. (The flight stabilized at 200 feet landed long.)

Three of the 11 flights involved challenging "slam dunk" clearances by ATC. In two other instances, ATC somewhat unusually left the management of approach speed up to the pilots, and the pilots did not slow down in time to stabilize the approach appropriately. More generally, several of

these approaches were unstabilized, in part, because the flying pilot did not manage drag and configuration optimally. Further, during no observed flight did a monitoring pilot make a specific callout of the nature and magnitude of a deviation from stabilized approach parameters (e.g., "we're twenty knots fast" or "we're two dots high on the glideslope"). These parameter callouts were not specifically mandated in the observed airlines' SOPs, so for the sake of consistency we did not record their absence as a deviation. However, see the Discussion section for our suggestion that it would be useful for airlines to require specific callout of the nature of some deviations.

4.4 Checklist and Procedure Design

Degani and Wiener (1993) reported that airline checklists were in some cases not well designed for effective performance. While our study did not address design of checklists and associated procedures, our general impression is that checklist design has improved considerably since the Degani and Wiener study. However we did note a few instances of problematic design of checklists and procedures that contributed to error vulnerability:

1. An Approach checklist that required suspending the checklist until an appropriate time to advise the flight attendants to be seated on final approach. (The airline revised this checklist to correct the problem during the course of our study.)

2. Prescribing reconciliation of final weight and balance numbers and flight management system (FMS) entries to be done just after pushback. This usually resulted in this task being performed at a poor time during pushback, engine start and initial taxi out on the congested ramp,

3. Requiring the monitoring pilot to make all FMS entries, overloading this pilot during descent and approach when he or she was busy with other tasks.

4. Prescribing an In-range checklist as a flow-then-check procedure. Although this is not necessarily poor design, we observed that this In-range checklist was actually performed as read-do in most cases. This suggests that the prescribed procedure may not work well in the actual operational environment.

4.5 Effective and Exemplary Monitoring and Checklist Performance

The deviations described in the previous sections constitute only a tiny portion of instances of checklist and monitoring execution. For context, we provide examples of the great bulk of instances of correct, sometimes exemplary execution. Crews routinely neutralized threats that, on other flights, have resulted in accidents. For example, as one observed flight was taxiing out of the gate, the ground controller issued instructions to another aircraft that would have caused a head-to-head conflict between the two aircraft. The first officer monitored the controller's mistaken instruction, immediately perceived the inherent conflict, and transmitted a challenge to the controller. Crews also frequently used checklists and monitoring to trap and neutralize their own errors, which otherwise might have undercut safety. For example, on one flight the captain failed to set his navigation radio to the instrument landing system frequency before or during his approach briefing at cruise altitude, as required by the FOM. Later, while the flight was being vectored in the terminal area, the Approach checklist item "Navigation radios – tuned and identified" helped the crew trap this error.

There were also many instances in which a pilot trapped a crewmember deviation by monitoring more extensively than specifically required by SOP. As one flight taxied onto the runway, the captain did not engage the autothrottle during the before-takeoff flow. The first officer noticed the captain's omission during the takeoff roll and turned on the autothrottle as the aircraft accelerated through 80 knots. As a result, the flight obtained proper takeoff power. On another flight, although

the auxiliary power unit (APU) was inoperative, the crew did not pre-brief the somewhat unusual situation of arriving at the gate with an inoperative APU. As the airplane approached the gate, the captain called for shutting down engine number 1, which was inappropriate with an inoperative APU. The first officer quickly pointed out that the APU was inoperative, and the captain decided to leave both engines running. On yet another flight, as the aircraft accelerated down the runway, the captain (the monitoring pilot) became preoccupied with a confusing power indication and omitted the airspeed callouts for V1 and Rotation. At about 10 knots faster than Vr the first officer noted his airspeed indicator and called out "Rotate." Simultaneously, he initiated rotation on his own.

Some situations were particularly difficult for a pilot to monitor and challenge for social/interpersonal reasons, yet in many of these the pilots spoke up about their concerns and trapped the deviations. For example, the captain omitted the pre-departure briefing on one observed flight. During the subsequent checklist, in response to the first officer's challenge of "Flight Attendant and Pilot Briefings," the captain responded, "Got any questions?" The first officer responded that he would like to have a briefing, so he briefed the departure. After this the captain added some relevant information, so he was induced to participate in the briefing, as well.

We also observed effective crew performance beyond checklists and monitoring. Many flights involved situational challenges ("threats"), and some crews were especially proactive in addressing these conditions. For example, one flight operated in weather conditions that were forecast to be marginal at the destination. The crew monitored the weather throughout the flight. During cruise, the captain anticipated the possibility of the weather deteriorating to low ceilings and visibilities, so the crew reviewed the company's "monitored approach" procedures and discussed the requirements. On another flight with an extremely short leg length and almost no level cruise segment, the captain slowed the aircraft down to provide adequate time to brief and prepare for a challenging approach that lay ahead.

We observed several examples of specific techniques that crewmembers used to enhance their performance, for example:

Deliberateness. One first officer had a nice technique of carefully pointing to the overhead panel items that he was calling out during the After Start checklist. This focused both pilots' attention on the checklist items and the specific switch settings/indications on the panel. After departure on another flight, a first officer set up the flight management system's climb page and then paused before executing to let the captain verify the change. The captain focused on the first officer's control display input (CDU) screen with apparent deliberateness to verify the change. The crew then performed this cross verification for every CDU input requiring execution for the remainder of the flight. On yet another flight, a captain wanted to initiate flowing the cockpit panels in the pre-departure phase, and he asked the first officer, "Are you watching me?" as a way of prompting the first officer to verify items being checked.

Modeling self-discipline and professionalism. On one flight, the captain interupted what he was saying in mid-sentence to closely attend to the aircraft and autopilot during level-off, thus demonstrating effective management of workload and attention. On another flight, the first officer attempted to initiate a brief non-essential conversation during taxi-out. The captain did not respond and the first officer did not continue the non-essential remarks.

Making an error-trapping routine more reliable. Many air carrier pilots use the aircraft's taxi light switch during final approach as a reminder that their flight has been cleared to land, turning the light on after receiving the clearance. However, unless a pilot checks the position of the light

switch before touchdown, this technique is not an effective safeguard against landing without clearance. One observed pilot, though, had formed a strong habit of checking the light switch at the same part of every approach (at 500 feet). This pairing of the reminder with habitual verification greatly improves the effectiveness of the technique.

5. DISCUSSION

Although checklists and monitoring are crucial defenses against threats and errors that might lead to accidents, these defenses sometimes fail. Our study was designed to provide data that can help the industry understand the nature of these failures and why they occur. Our approach was to conduct cockpit jumpseat observations from a cross-section of airline flight operations conducted in six often-used aircraft types. Data were collected from three well-established, relatively stable airlines (only two observations were conducted at one of the airlines because of practical constraints). Although deviations in primary procedures used to operate the aircraft were not a central focus of this study, data on these deviations were also collected to determine the extent to which checklists and monitoring trap deviations.

The average number of deviations observed per flight was fairly large: 3.2 for checklists, 5.2 for primary procedures, and 6.5 for monitoring (15.0 total). Few studies have been published with which to compare these deviation rates. The Degani and Wiener study of checklists was qualitative rather than quantitative. Although a large amount of LOSA data has now been collected in the airline industry, few quantitative studies have been published. Klinect, Wilhelm, and Helmreich (1999) reported an average of 1.84 total errors per flight in 314 LOSA flight observations at three airlines, and presumably they defined errors in the same way we defined deviations. Klinect et al. reported that 64% of flight segments had at least one error, whereas all of our flights had at least one deviation. (However, 32% of our flights had no checklist deviations and 13% had no primary procedure deviations.) Our substantially higher rates presumably reflect the fact that we deliberately attempted to record as many deviations as possible, even minor ones with no apparent consequences; also, focusing on the details of checklist use and monitoring allowed more events to be recorded on these aspects than is practical in typical LOSA observations.

Deviations in checklist use, monitoring, and primary procedures undoubtedly occur for diverse reasons, discussed in considerable detail later. Whether these deviations constitute a problem for safety depends on the specific deviation and the circumstances under which it occurred. In some cases deviations were undoubtedly driven by competing operational demands and were appropriate. For example, an ATC radio message might be received at the time a pilot would normally make a "thousand feet to go" call, and it is appropriate to focus on this message and then turn attention to monitoring for level-off, even though this results in the call being made after the chime. In other cases, deviations were almost certainly inadvertent and unwitting, and these can properly be called errors. Other deviations may reflect poor habits in complying with SOPs, and these fit the definition of "violations" (Klinect et al., 1999). Still others may reflect deviations so widespread through a pilot group that they become norms for line operations; when this is the case it is necessary to analyze why deviation has become normalized, and in some cases the procedure should be redesigned. Our data provide a basis for understanding how operational demands, human cognitive constraints, and organizational factors affect the ideal execution of procedures prescribed in FOMs.

As we stated in the introduction, the number of deviations per flight should be considered in the context of the number of opportunities for deviation. For example, one airline used 10 checklists with a total of 197 challenge items plus response items. Several types of deviation could be made

for each item (failure to respond, using non-standard phraseology, failure to look at item checked, etc). Thus, even if we considered all of these deviations to be errors, the rate of occurrence in terms of errors per opportunity was probably well under one percent, which is in the ballpark for many forms of skilled human performance. Put another way, in the vast majority of cases, checklists and monitoring were performed appropriately. Error rates, of course, vary enormously as a function of the nature of the task and the conditions under which it is performed (Reason, 1990), and later on we discuss the diverse factors that probably influenced each type of deviation.

The number of checklist deviations in each phase of flight was roughly proportional to the number of checklist items performed in that phase, with the most deviations occurring in pre-taxi, taxi-out, descent, and approach phases. In contrast, monitoring deviations were especially associated with climb and descent, and primary procedure deviations were mostly distributed among pre-taxi, taxi-out, climb, cruise, descent, and approach (Tables 3 and 4). Beyond the number of opportunities for each category of deviation in each phase of flight, other task demands occurring concurrently probably contributed to vulnerability to deviation (Loukopoulos, Dismukes, & Barshi, 2009). The distribution of deviations among phases of flight was similar to that reported by Klinect et al. (1999).

Variability among flights was quite large, ranging from one to 38 total deviations per flight. Distribution of deviations per flight was skewed substantially to the right. For example, on 31 flights only 0-2 checklist deviations were observed, but the remaining 29 flights ranged from 3 to 13 checklist deviations. Several factors may have contributed to this variability: 1) imperfect standardization of performance among pilots (i.e., variability between pilots); 2) random variation within pilots from one flight to the next; 3) variations in the demands and conditions between flights; 4) random variation in the observer's noticing deviations; and 5) differences in the observer's level of familiarity with different aircraft types and company procedures.

Our data do not allow us to determine which of these factors were at play, but we suspect all five played a role. For example, the number of deviations made before takeoff was moderately correlated with the number of deviations after takeoff in each category, suggesting some consistency within particular crews in making more or fewer deviations. Also, the substantial clustering of a large subset of the crews making few deviations suggests that this subset followed procedures relatively well in comparison to the remainder who varied greatly in the number of deviations committed. Although the observer is an airline pilot highly experienced in several aircraft types, and he carefully studied each FOM before observing, observation was probably more nuanced for the aircraft and specific procedures with which he had the most experience. The observer identified more primary procedure and checklists deviations, but not monitoring deviations, in the aircraft and SOPs with which he was most familiar. (Data not shown.) Also he was sometimes quite busy taking notes during some phases of flight in which the crews had many tasks to perform, and this undoubtedly affected what he was able to detect and record.

5.1 Types and Possible Causes of Deviation

Rather than creating a deviation taxonomy *a priori*, or using one of the several error taxonomies that have been proposed for cockpit operations, we sorted each of the three deviation categories (checklist, monitoring, and primary procedure) into types according to similarity in operational aspects. Checklist deviations clustered into six types: flow-check performed as read-do; responding without looking; checklist item omitted, performed incorrectly, or performed incompletely; poor timing of checklist initiation; checklist performed from memory; and failure to initiate checklists (in

order of number of occurrences; Table 5). The first two types accounted for nearly half of the checklist deviations observed.

Diverse factors undoubtedly contributed to these checklist deviations. For example, one checklist duplicated almost all the items on the preceding flow, rather than covering just the most safety-critical items, and pilots may have found it more straightforward to perform this checklist as a read-do than as a flow-check. (One manufacturer's philosophy for its recent generation of aircraft is to minimize the number of items on checklists, thus discouraging converting flow-check procedures to read-do.) Operational demands may also have come into play. Crews are at times under considerable time pressure, with many tasks to perform, and may sometimes perform a flow-check procedure as read-do to save time.

Responding without looking may reflect two quite different situations. In some cases pilots may respond from the memory of having set or checked the item only moments before as part of the flow. Written procedures are often vague about whether the pilot is expected to visually re-check items set/checked during the flow or just respond from memory, and some items must be checked through recall from memory (e.g., whether the ground crew has signaled that ground equipment has been removed). Responding from memory reduces the intended protective redundancy of flow-check procedures; also, pilots may not be aware that they are vulnerable to source memory confusion (Dismukes, Berman, & Loukopoulos, 2007, p. 113), in which memory of the current situation is confused with memory of instances of having set/checked an item on many previous flights (in some cases the most recent previous flight may have been only a few hours before). Pilots may respond from memory habitually or only when under time pressure.

In some cases what we recorded as responding without looking may actually have been instances of "looking without seeing". Expectation that an item is correctly set arises from memory of having just set or checked an item and from the vast number of previous instances in which that item has been correctly set. Thus, even though the pilot may direct gaze toward the item to be checked, he or she may perceive it to be in the correct position even when it is not, especially if gaze fixation on the item is brief due to rushing. Also, it is possible that pilots' response to the checklist challenge may become so automatic that pilots sometimes utter the response automatically, perhaps not even realizing that they have not visually confirmed the challenged item.

Performing checklists from memory is a clear violation of formal procedures, but airlines may underestimate the factors that encourage pilots to do this. In general, when an individual has performed a simple task, such as executing a checklist, many times, performance becomes largely automatic, fast, and fluid, requiring little cognitive effort. To force oneself to read an often-performed checklist by reading each item feels cumbersome and effortful and slows down execution—often at a time when the crew is hurrying to complete cockpit preparations. This analysis does not excuse the deviation from formal procedures, but does suggest that airlines must make clear that they expect crews to slow down and take the deliberate and effortful approach of reading checklists item by item. This raises the issue of whether the airline industry is giving conflicting messages to pilots: slow down and be deliberate, but respond quickly to frequent time pressures.

Failure to initiate checklists, at least in large airline operations, almost certainly reflects memory failures, probably as a result of distractions and other competing demands on pilots' attention, or of circumstances forcing procedures to performed out of the normal sequence (Loukopoulos et al.,

2009). In contrast, initiating checklists at a poor time (e.g., when both pilots should be attending to more urgent tasks) probably reflects poor task management. Pilots typically do not receive detailed training on timing of checklists in the context of competing task demands, and this is an appropriate topic for recurrent training.

Monitoring deviations grouped in three clusters: late or omitted callouts, omitted verification, and not monitoring aircraft state or position (Table 6). Over half of the monitoring deviations were late/omitted callouts, most of which (140) were the "1000 feet to go" call, required as the aircraft approaches level-out altitude. Often the "1000 feet to go" call was prompted by the altitude chime at 1000 feet, which removes the redundant protection designed into the procedure; however, it is not surprising that this callout is often late. It must be made in a very short time window (a few seconds), which requires close monitoring of the altimeter at a time when pilots often must divide attention to monitor other tasks. Also, automation complacency may creep in, because the chime is highly reliable. The danger, of course, is that the automation on occasion has not been set properly and the chime does not sound because the automation is not preparing to level the aircraft. We estimate that the "1000 feet to go" callout was missed around 1/3 of the time, which raises the question of the effectiveness of this callout and whether it should be revised in some way. Simply exhorting pilots to make the callout as prescribed is not likely to change performance substantially. One approach, which would require study, might be to accept that this call will fairly frequently occur only after prompting by the chime (which accomplishes the purpose of alerting the crew to start monitoring for level-off), and focus on finding ways to reduce deviations in setting the target altitude in the flight management system and to help pilots better recognize implications of automatic mode changes.

A much more serious omission was failure to make callouts required during unstabilized approaches. When the flying pilot is trying to get a "slam dunk" approach[9] stabilized, the pilot may be so busy that he or she may not recognize how far the aircraft is from acceptable parameters and urgently needs prompting from the monitoring pilot. It is not clear why unstabilized approach callouts are sometimes omitted. Monitoring pilots may erroneously think the flying pilot fully recognizes the extent of the situation or may think that saying something may distract an already overloaded flying pilot. Failure of the monitoring pilot to verbally challenge an unstabilized approach removes the opportunity to alert the flying pilot to the nature of the situation and to prompt the correct response, which is to execute a go-around. The NTSB has found that omission of these challenges has contributed to several landing accidents (Dismukes et al., 2007, Chapters 5 and 19). The importance of making these callouts and the rationale may not be sufficiently emphasized in training and checking.

In other situations the monitoring pilot's decision may be more reasoned. For example, at the point at which a go-around from an unstablized approach is prescribed by SOP (typically 500 to 1000 feet above ground), the flying pilot may have managed to have gotten the aircraft properly configured for landing, on glideslope and localizer, with airspeed and descent rate on target but not yet have the engines spooled up as required. Technically the monitoring pilot at this point (according to some airline SOPs) should call "Unstable, go-around" but might choose not to do so seeing that the flying pilot recognized the situation and was about to advance the throttles. This example illustrates a

---

[9] These are approaches in which ATC puts the aircraft in a position too high and too fast for the crew to easily configure the aircraft for landing, to capture glideslope, and establish the appropriate airspeed.

difficult tension between writing SOPs to cover critical situations and allowing pilots to exercise reasonable judgment. On the one hand, if SOP only recommends—instead of mandates—going around at this point, some pilots may use the latitude to continue unstabilized approaches far too close to the ground. On the other hand, if airlines officially state that pilots are not to deviate at all from unstable approach criteria, but at the same time strongly encourage them to pursue on-time performance aggressively, and if the airlines do not provide realistic guidance on how to resolve these competing objectives, pilots are likely to conclude that SOP is pro forma. This "wink and a nod" stance may lead to widespread deviation from all SOP for pilot convenience or company profit. Thus, if companies intend a "bottom line" to be adhered to without even small variation, this must be strongly emphasized in training and in line checks. If pilots are allowed to exercise judgment about small deviations in specific situations, the limits of deviation should be discussed explicitly.

At the observed airlines, monitoring pilots were not required to specify the nature and magnitude of deviations (e.g., "20 knots fast"); however, airlines may want to consider requiring callouts of specific parameters of deviation. Parameter-specific deviation callouts can be highly effective at providing or restoring situational awareness to a flying pilot who might be distracted or overloaded with sensory inputs during an unstabilized approach. A specific deviation callout by the monitoring pilot early enough in an approach can help the flying pilot stabilize the approach before reaching the bottom line altitude for stability; in contrast, at the bottom line a go-around is mandatory regardless of the nature of the deviation, so calling out the nature of the deviation may be less relevant than earlier in the approach. Further, requiring parameter-specific callouts may make it easier for the monitoring pilot—especially if a first officer—to frame a challenge.

Diverse factors may contribute to omission of required verifications, such as: 1) SOPs that do not specifically define what is to be verified; 2) SOPs that combine multiple verifications into a single checklist item, making it easy to omit one or more verifications; 3) the human tendency to "look without seeing" when performing a routine repetitive task; and 4) the challenge of deliberately pacing verification in a fully conscious manner when under time pressure. Airlines should emphasize more strongly the need and rationale for slow, deliberate verification of items, and explain subtle cognitive factors that undercut performance. For example, because items to be verified are almost always in the expected position (e.g., the three green landing gear down position indicator lights come on after the gear handle is selected for down), pilots are subject to expectation bias. Also, they may not be aware that their verification habits have slowly eroded over time, because no feedback occurs when verification is not done carefully—usually nothing happens when verification is not done because the item to be verified is set appropriately. (See Dismukes et al., 2007, for an extended discussion of this phenomenon.) These factors also apply to checklist execution as well as to effective verification.

The third type of monitoring deviation, failure to monitor aircraft state or position, could in some situations seriously undermine safety. Some instances of this deviation probably resulted from competing concurrent task demands on attention. Human ability to divide attention among tasks is quite limited, usually accomplished by switching attention back and forth, which leaves individuals vulnerable to losing track of the status of one task while engaged in another (see Loukopoulos et al., 2009 for an extended discussion of this problem). Although Crew Resource Management (CRM) classes include a module on workload management, these modules typically focus on prioritization and distribution of workload among crew members, which are important topics, but no guidance is provided for how to manage attention when juggling concurrent task demands.

Effective monitoring is far more difficult to maintain than may be apparent, especially when human operators are not directly controlling the system being monitored. Equipment failures are infrequent in modern airline operations, and humans are inherently poor at monitoring for infrequent events. Pilots do not receive feedback on the effectiveness of their monitoring, comparable to the immediate feedback the aircraft gives them if they mishandle the controls when flying manually; consequently, pilots are not likely to recognize that their monitoring in inconsistent.

Although this study focused mainly on checklist use and monitoring deviations, the additional data on primary procedure deviations provide context and allowed us to examine how effective checklists and monitoring were at trapping primary procedure errors. We grouped the 15 types of primary procedure deviations into six areas: 1) coordination within the crew or with ATC; 2) use of automation; 3) approach stabilization; 4) path and airspeed control; (5) configuration of systems or flight controls; and 6) planning and execution (Table 7). By far the most common deviations were failure to properly configure systems (62 instances), poor planning for contingencies (57 instances), poor coordination between the pilots (56 instances), and problematic use of the FMS (40 instances). Most of these deviations appeared to be inadvertent and can properly be described as errors.

System configuration errors, when not caught by monitoring or checklist use, can have serious consequences, as illustrated by the Helios accident described at the beginning of this report. These errors, as well as some of the other deviation types, are almost certainly slips or oversights, that may result from competing task demands or poor procedure habits. The 40 FMS errors illustrate that problems with automation design, cockpit interfaces, and training—noted from the first introduction of FMSs—continue in spite of numerous studies (e.g., Sarter & Woods, 1994; Sarter, Mumaw, & Wickens, 2007). Contingency planning and execution shortcomings and poor coordination between pilots and other personnel are CRM failures. Although CRM has become widely accepted since its inception 20-odd years ago and is taught at most airlines, these deviations suggest that much room remains for improvement. We fear that CRM training and checking have become somewhat pro forma and receive less emphasis in this era of drastic cost-cutting in the airline industry. Particularly disquieting is the low percentage of deviations crews detected and trapped (discussed later in this section.)

It is difficult to compare the distribution of crew deviations we observed among three categories and 24 types (sub-categories) with error distributions others have reported, in part because not all of the deviations we recorded should be considered errors. Also, we deliberately avoided creating categories *a priori*, rather, for the purposes of our study, grouped deviations *post facto* by the operational action involved. Other studies have used very different categories/types, some of which are descriptive and some of which are based on *a priori* theoretical distinctions (e.g., Klinect et al., 1999, Sarter & Alexander, 2000; Thomas, 2004). It would also be useful to compare our distribution with error types identified in airline accidents, even though the sampling is profoundly different—our observations involved non-accident flights. Unfortunately, analyses of pilot error in airline accidents have also used very different categories/types (Lautmann & Gallimore, 1987; NTSB, 1994; Shappell, Detwiler, Holcomb, Hackworth, Boquet, & Wiegmann, 2007;. Li, Grabowski, Baker, & Rebok, 2006), making numerical comparison almost impossible because we do not know in what manner the deviations we observed would be classified under these error taxonomies. However, in a later section we do discuss similarities and differences between our observations and accident report findings about pilot error.

Klinect et al. (1999) reported that the largest (54%) of their five categories of error in LOSA observations was intentional noncompliance. (The other four were procedural, communication, proficiency, and operational decision.) Although some of the deviations we observed were clearly conscious noncompliance—performing a checklist from memory, for example—it would not be possible to determine whether many of the deviations were deliberate or unwitting. A captain's failure to call for the Before Takeoff checklist might occur either because he or she had the inappropriate habit of allowing the first officer to self-initiate that checklist or because distraction diverted attention from making the intended call. Some deviations are clearly unintentional, such as deviations from flight path.

5.2 Factors Affecting Deviation

In an analysis of 12 years of airline accidents attributed to crew error, the NTSB (1994) found that 55% were running late, considerably higher than their sample of non-accident flights. In contrast, we did not find any indication that crews on late flights made more deviations. However, this difference is consistent with our previous finding (Dismukes et al., 2007) that accidents attributed to crew error very rarely are the product only of a single error; instead these accidents typically result from the convergence task demands, happenstance events, organizational factors, and human factors.

The crew pairing procedures of large airlines result in two pilots often flying together for the first time at the beginning of their trip. The NTSB (1994) found that accident crews were often on their first flight or first day of flying together. In our study, pilots who were on their first flight together (14%) or on their first day together (32%) made substantially more monitoring, checklist, and primary procedure deviations than those crews not flying together for the first time; however, this difference was statistically significant only for primary procedure deviations. (Our small sample size provided limited statistical power.) Consistent with previous studies (Foushee, Lauber, Baetge, & Acomb, 1986; Thomas & Petrelli, 2006), these results suggest that as two pilots fly together they settle into a more effective mode of working together. In particular, their actions may become more coordinated, and they may become more comfortable challenging deviations the other pilot makes.

To what extent does experience affect deviation probabilities? On average, captains typically have more overall flight experience than first officers, however we found that captains and first officers were equally likely to make deviations in both the flying pilot role and the monitoring pilot role. Pilots (captains and first officers combined) in their first year in their aircraft type and in their role (captain or first officer) did not make more deviations than pilots with more experience in their position and aircraft type. A separate analysis of just first officers found that those in their first year in position and type did not make more deviations than more experienced first officers. Caution should be used in extrapolating this finding to other airlines. The two airlines from which most of our data came require high levels of experience for newly hired first officers. It would be useful to repeat this study with regional airlines which are hiring first officers with only a few hundred hours of flight time and who are new to airline operations. However, our finding with two major airlines should be reassuring that captains and first officers in their first year in aircraft type and position are not slow to reach proficiency.

5.3 Deviation Trapping

Only 18% of deviations—even those that were clearly errors—were trapped (caught and corrected) or even discussed, a disquieting finding. In comparison, Klinect et al. (1999) reported that 36% of errors observed in LOSA were trapped, and Thomas and Petrilli (2006) reported 63% were detected and actively managed in a flight simulation study. Our lower trapping rates probably reflect

multiple factors, one of which is that we observed actual line operations, in which operational pressures and opportunities for error are not fully captured by simulations. Also, the lower trapping rate we observed may reflect the fact that we deliberately recorded even very minor deviations, which is probably not true of most LOSAs. The percent of deviations trapped varied greatly across deviation types. In general, primary procedure deviations were more often caught: 35% versus 14% of checklist deviations and 6% of monitoring deviations. It is not surprising that monitoring deviations were least likely to be caught, since monitoring can be considered a final defense against primary errors (Sumwalt, et al, 2002). Very large differences in trapping occurred among the types of deviation within each category. Only one of 113 verification omissions, 12 of 211 late or omitted callouts, and one of 48 flow-checks performed as read-do were trapped. In contrast, 25 of 33 failures of crew-ATC coordination, 14 of 18 MCP deviations, and 32 of 62 system configuration deviations were trapped.

These large differences in trapping of different deviation types may reflect how conspicuous the consequences of the deviation are to the pilots and other personnel. Also, whether one pilot challenges a deviation by the other pilot may reflect how dangerous the deviation is perceived to be. In some situations, even when one pilot detects the other's deviation, it may be difficult or awkward to challenge the deviation. For example, "one thousand to go" calls must be made shortly before the altitude alerter chimes, and it is not clear to the flying pilot until the chime sounds whether the monitoring pilot will make the call. (At some airlines, the flying pilot makes this callout.) Further, the monitoring pilot—especially if a first officer—must consider whether frequently pointing out deviations that are unlikely to be consequential will create a tense cockpit. Similarly, a captain must be selective about challenging errors made by the first officer in order to avoid micromanaging the flight deck, which undercuts open communication.[10] On the other hand, in some situations it is difficult for a pilot to assess in real time whether an error will have significant consequences. Any missed callout or verification removes the power of that action to trap errors and prevent undesired aircraft states.

Captains in the monitoring pilot role were more than twice as likely to trap deviations made by the flying pilot than first officers in the monitoring pilot role (27.9% vs. 12.1%). This is consistent with flight simulation research showing that captains were more likely to challenge first officers flying the aircraft than vice versa (Orasanu, McDonnell, & Davidson, 1999; Fischer & Orasanu, 2000) and is also consistent with the 1994 NTSB study of accidents attributed to crew error. The simulation studies also revealed that captains were more likely to use commands and first officers to use hints to call the flying pilot's attention to errors, high risk errors were more likely to be challenged than low risk errors, and first officers were less likely to challenge an error if the error involved a loss of "face" for the captain. The most common reason both captains and first officers gave for not challenging an error was that they noticed the error but felt that no intervention was necessary—the deviation was minor. The next most common reason was that they had not noticed the error, indicating failure in monitoring.

Interestingly, we found that when captains and first officers were the flying pilot they were about equally unlikely to challenge deviations made by the monitoring pilot (7.3% vs. 9.5%). The low deviation-trapping rate in the flying pilot role may reflect both that the flying pilot was too busy to catch monitoring pilot deviations and that pilots gave low priority to this.

---

[10] We are indebted to a senior airline captain for pointing this out.

Our data, taken together with the flight simulation studies, indicate that some deviations are simply not noticed and that when first officers notice deviations they are less comfortable challenging them. The Orasanu et al. (1999) finding that both captains and first officers find some—perhaps many— errors not important enough to challenge, and perhaps embarrass the other pilot, raises questions about how realistic are industry expectations about the monitoring pilot role. Should the monitoring pilot challenge even small deviations very likely to be inconsequential, possibly at the risk of undercutting cockpit harmony and being perceived as nitpicking? How should first officers go about challenging so they can be trap deviations at least as frequently and assertively as captains, and what kind of support do first officers need from company procedures, training, and culture to be able to challenge effectively?

The industry needs research to find out why, overall, deviations are trapped so infrequently. In the interim, based on what is already known, the industry should further emphasize the importance of challenging and should provide specific guidance, training, and practice on how to challenge.

5.4 Outcome of Deviations

Based on the sample of slightly more than half of the flights that we evaluated as to consequences, eighty-nine percent of the observed deviations had no discernable outcome other than an arguably small reduction in the efficacy of safeguards. For example, even though pilots sometimes failed to make the "thousand feet to go" call the autopilot leveled the aircraft at the correct altitude, though of course if the FMS or MCP had been set up incorrectly, the aircraft might not have leveled off. The fact that the great majority of deviations do not lead to serious consequences suggests that the overall system of multiple, overlapping safeguards works fairly well. However, nine percent of deviations led to an undesired aircraft state, and two percent led to subsequent deviations. In comparison, Klinect et al. (1999) reported 85% of LOSA errors were inconsequential, 12% resulted to an undesired aircraft state, and 3% in addition errors. (This suggests that Kilinect et al. used the term error in much the way we use the term deviation.)

We observed 45 instances of undesired aircraft state of diverse sorts: deviations in airspeed, heading, or vertical path; incorrect heading set for takeoff; incorrect configuration of controls or systems; flight attendants not seated when required by SOP; unstabilized approaches and landing from unstabilized approaches; inadequate terrain separation, etc. (Table 12). Clearly these undesired states—some resulting from multiple deviations—were more serious than the outcome of most deviations in that the potential for an accident was greater.

The most common undesired state was mis-configuration of aircraft systems, typically resulting from failing to set a switch correctly during a flow, for example failing to turn on windshield heat or failing to set cockpit/cabin pressurization properly. Some of the mis-set items were items on checklists—in these cases the item was missed on both the flow and the checklist. Because the number of items that have to be set and/or checked on each flight is large, opportunities for missing an item abound. (However, more modern airliners have fewer items that have to be set on each flight.) Even skilled, conscientious pilots are vulnerable to not perceiving that an item is not set correctly, for cognitive and operational reasons already discussed in the context of failures of monitoring and checklist use. The potential danger of these errors is illustrated by the Helios accident, in which the chain of events leading to the accident began with failing to identify an incorrect setting of the pressurization panel. Thus training, checking, and the design of cockpit procedures should be bolstered to address this crucial vulnerability. (See the later section on Countermeasures.)

Particularly troubling were unstabilized approaches in which mandatory callouts were not made. Eleven of the 60 approaches were unstabilized at some point according to the respective airline's published criteria. In two of these 11 approaches, the crew was able to stabilize the approach before the final gate (1000 feet or 500 feet, depending on the airline); one crew appropriately executed a go-around, but eight crews continued to land. Five of these eight were stabilized by 500 feet, which raises a question of whether a 1000 foot final gate is completely realistic, and—if it is realistic—whether rigorous compliance is adequately emphasized.

Attempting to land from an unstabilized approach has played a central role in several major airline disasters in recent years (Dismukes et al., 2007). The pernicious threat of continuing unstabilized approaches beyond the final gate is that trying to get the aircraft parameters into acceptable limits requires so much of the crew's attention that they may not be able to judge whether they will succeed in making the approach and landing work out. Failure of monitoring pilots to make required callouts of deviations contributes to the problem. Even though many airlines have now appropriately adopted no fault go-around policies, the industry may not understand how strongly both operational pressures—such as strong emphasis on saving fuel—and inherent cognitive processes push pilots to continue unstable approaches. Both organizational and cognitive processes are insidious because they sometimes operate unconsciously. Pilots may not always be aware that their decisions are affected by concerns with on-time performance and fuel costs, and they may not recognize that having landed from an unstabilized approach several times at long runways is skewing their judgment about risks that become all too apparent when the runway is short.

Should crews be given some latitude in deciding whether passing through the gate without all parameters (configuration, airspeed, glideslope, localizer, and engines spooled) on target requires an automatic go-around if they judge they will quickly be on target? In any case, what is counterproductive and inappropriate is for companies to formally require exact adherence to stabilized approach criteria, yet implicitly expect those criteria to be bent in practice to satisfy production pressures.

5.5 Accidents and Normal Flights

One objective of this study was to compare the kinds of deviations observed in normal flights with the errors uncovered in accident investigations. Most aviation accidents are attributed, at least in part, to pilot error. What determines whether errors and other deviations lead to accidents? Answering this question could support developing more effective ways to prevent accidents. We focused on checklist use and monitoring because these are two of the major defenses designed to detect threats and errors and keep them from escalating into accidents.

Many of the types of deviation reported here have over the years contributed to airline accidents; for example, missed items on checklists, erroneous inputs to automation, failure to recognize and plan for downstream implications of evolving situations, unstabilized approaches, and—most serious—failure to monitor and challenge errors. Our finding that first officers in the monitoring pilot role were substantially less likely to trap the flying pilots' deviations than were captains in that role is consistent both with simulation studies and the NTSB's (1994) analysis of accident factors. However, even the deviation trapping by captains was too infrequent to provide the level of protection that the industry seems to assume.

A few of the un-trapped deviations we observed led to undesired aircraft states, which increase the risk of an accident, however, the vast majority of deviations had no observable outcome. One statistical factor does distinguish the 60 flights observed here from many accident flights. Dismukes et al., (2007) found that a surprising large percentage of major accidents occurred when crews had to respond very quickly to sudden threats such as windshear, false stall warnings shortly after takeoff rotation, or loss of flight instruments. In contrast we observed no such situations demanding fast response. Although these situations are extremely rare, when they do occur, pilots are severely challenged to diagnose the situation and choose the best response. Thus, one avenue to preventing accidents would be to identify scenarios representative of diverse sudden-threat situations and to devise systems, procedures, and training to help pilots respond effectively.

Beyond the sudden-threat issue, the deviations and situation contexts for those deviations we observed are remarkably similar to the errors uncovered in accident investigations, so what determines when deviations are inconsequential and when they lead to accidents? We suggest three factors:

1. The aviation system has many layers of protection. Thus, for example, when a crew fails to monitor the aircraft leveling off under automation, the vast majority of the time the automation has been correctly programmed and levels off correctly. Even if the aircraft does not level off, the air traffic controller may notice and call this out to the crew or TCAS (traffic collision avoidance system) may provide last-minute traffic separation, but of course there is a very small but finite chance that failing to level out will lead to a mid-air collision.

2. The great majority of the accidents analyzed by Dismukes et al., (2007) resulted from the somewhat random co-occurrence of multiple threats and errors. These multiple problems combined in a more than additive fashion, making the situation far more difficult to manage.

3. As several factors and errors coincided in these accidents, the challenges of the situation and the crew's workload snowballed, sometimes overwhelming the crew. Monitoring and error trapping often fell by the wayside as the crew fell behind in the face of mounting task demands. In contrast, our current study shows that in the vast majority of flights multiple factors do not combine to overwhelm the crew.

The latter two factors pose a major challenge to efforts to improve aviation safety, because the industry can create barriers to individual threats and errors, but the number of possible combinations of multiple threats and multiple errors is astronomical. This challenge underscores the importance of determining why error-trapping rates are low and developing better ways to train and support error trapping. Training should focus especially on helping pilots recognize the particular danger of multiple threats—each of which may be by itself managed with routine precautions—and on techniques for backing out of situations with escalating workload.

6. COUNTERMEASURES

People sometimes assume adhering to checklist and monitoring procedures is simply a matter of pilots being disciplined and professional. Although discipline and professionalism are essential, they are not sufficient to achieve the level of performance required, because of the cognitive, task, and organizational factors discussed earlier. We suggest a range of countermeasures that could reduce pilots' vulnerability to the deviations observed in this study.

Some of the deviations from SOP we observed were probably intentional, in some sense, but we think most of these deviations were unwitting. In either case, simply exhorting pilots to follow

procedures exactly as written will have limited effect—it is necessary to understand the factors that lead to both intentional and unintentional deviations and develop specific countermeasures directed to those factors. Vulnerability to the types of deviation described in this report can be reduced by thoughtful design of training, checking procedures, operating procedures, organizational policy and practices, and system design. Existing knowledge is sufficient, if applied properly, to accomplish much in each of these areas, and research can provide still greater progress.

6.1 Cockpit Procedures and Organization Policies

Suggestion: Formalize monitoring and challenging requirements and procedures. Recognizing the importance of monitoring, some airlines have changed the designation of "pilot not flying" to "monitoring pilot". This is a good first step, but a detailed description of what is to be monitored and how it is to be accomplished is crucial for compliance, especially since humans are inherently poor monitors. Specifying callouts to be made in specific situations addresses some issues. For example, in recent years airlines have begun to formally prescribe the call outs monitoring pilots should make during approach and these callouts can help both pilots keep track of whether the approach is stabilized and can lead them to the appropriate response. Explicitly defined callouts make it easier to know when and how to challenge the flying pilot. Besco (1995) advocated escalating callouts to alert the flying pilot to deviations: probing, alerting, challenging, and—if at all else fails—emergency warning. Industry representatives and the research community should collaborate to develop best practices for monitoring and challenging.

Suggestion: Minimize checklist items involving multiple components and specify responses for each component. We noted frequently that pilots were incompletely performing checklist items with several steps—for example, the challenge "Hydraulics—Checked and On" was intended to direct pilots to look at both the hydraulic gauges on the forward panel and the hydraulic pump switches on the overhead panel, but some pilots checked only the overhead panel. Eliminate multiple-step response items if practical, and if not, require each step of the response to be stated. For example, the Hydraulics response might be "Gauges checked and switches on". Some other checklist items use a single challenge that is supposed to generate verification of multiple switches/indicators followed by a single response to stand for all. For example, one challenge-response element of an After Takeoff checklist was "Pressurization"—"Checked." The airline's detailed SOP for this element required verification of four switches controlling engine bleed and air conditioning packs as well as the indicators for cabin altitude and pressure differential. During the course of this study, the airline amended this checklist item to read "Bleeds and Packs"—"On and Auto." This revised item improved specificity about what was to be checked but, in our view, did not adequately direct pilots' attention to all of the required verifications. An alternative might be to have two checklist items:

1. "Bleeds and Packs" with a response of "Bleeds On, Packs Auto"
2. "Pressurization" with a response of "Auto"/"Standby" (as appropriate) or "Differential 2.0" (or whatever number the indicator showed).

However, this increases the length of the checklist, which increases the risk of missing an item, so the trade-off would have to be considered.

Suggestion: Evaluate error vulnerability of existing procedures and strengthen them. Procedures such as "point and shoot" focus both pilots' attention on the task performed and reduce vulnerability to "looking without seeing" error. In the "point and shoot" procedure one pilot points to a new entry in the altitude selector and the other pilot verbally confirms the entry and, at some airlines, also points to the display. This example illustrates a general principle, which is especially important for

checklist use: Execution should always be deliberate and not rushed, so that the executive portion of the brain is able to track and oversee the largely automatic operation of highly practiced actions. Gaps exist in knowledge of the best way to design checklists (both normal and non-normal) for specific aspects of operation. Existing checklists have evolved largely through trial and error, but few studies have been conducted to validate the assumptions underlying the design of these checklist; thus empirical research is needed.

Suggestion: Organizations should periodically review cockpit operating procedures to identify and relieve "hotspots" in which prospective memory and concurrent task demands are high and interruptions are frequent. Checklists and related operating procedures, created to operate cockpit systems appropriately, may not be well designed to be performed in the sometimes hectic operational environment. A careful analysis of the actual (rather than ideal) operational environment may suggest ways to improve the timing and structure of checklists to reduce competing task demands and distractions (Loukopoulos et al., 2009). Airlines can use ASAP (Aviation Safety Action Program), LOSA, and FOQA (Flight Operations Quality Assurance) data to identify specific parts of the normal SOP and even specific routings and locations at which pilots are frequently rushed in particular procedures and then revise procedures and provide guidance to relieve the pressure to rush.

Suggestion: Organizations should systematically analyze the entire body of explicit and implicit messages given their pilot corps to balance competing goals. Consciously or unconsciously, pilots may allow concern with on-time performance to rush execution of checklists and short-change monitoring, and airlines may, deliberately or not, over-emphasize this concern. Because rushing substantially increases error rates, airlines should carefully examine the trade-offs of policies such as reducing time allowed for turns (the time between landing and pushing back for the next leg of the trip). Also, because of severe economic conditions, airlines now strongly emphasize reducing fuel use and/or fuel upload, and this can influence pilots' decision-making in unintended ways.

Pilots notice what is being evaluated by their companies during check rides and line checks, and what is not; if proper checklist use or unstabilized approach call-outs are not strongly emphasized, pilots perceive these to be less important than getting the airplane on the ground on time. As part of this analysis, organizations should evaluate how realistic are their formal "bottom-line" requirements (e.g., executing a missed approach if the aircraft is not stabilized at a specified altitude on approach) in the light of actual line operations. If these requirements are too idealistic or too conservative, they should be modified. If the organization truly requires the prescribed actions without exception, these actions must be strongly reinforced in training and checking, and the reasons exceptions cannot be made must be explained clearly. The worst of situations is a "bottom line" that is routinely violated, for whatever reason—this promotes normalization of deviance across all areas of operation.

Beyond checking, LOSA, ASAP, and FOQA provide data reflecting on how well checklists and monitoring are being performed. Feedback to the pilot corps from all these sources of data should include a frank and realistic discussion of company expectations on balancing competing goals.

Suggestion: Organizations should examine the role of organizational procedures in vulnerability to error in the cockpit (as well as errors in the cabin, dispatch center, and maintenance hangar). For example, single-engine taxi, quick turns, and distribution of SOP revisions to pilots by memo can increase vulnerability to error. It is important to explicitly analyze the trade-off between increased error rates and efficiency and cost, rather than assuming that no downside will occur. Pragmatically,

the current economic climate of the airline industry makes it difficult for any one company to operate in ways that drive its costs above that of its competitors, which suggests that analyzing and balancing the trade-offs requires industry-wide effort.

6.2 Training, Checking, and Mentoring

Initial (new-hire) training and transition to new aircraft-type training focus primarily on teaching pilots aircraft-specific and company-specific operating procedures. Checklist use and monitoring procedures are included, and during this training pilots gain initial levels of proficiency in operating the aircraft in the prescribed manner. In many airlines this training also includes a module on Crew Resource Management (CRM), which may address error management and which may touch on broader issues of human factors. However, although pilots are often exhorted to follow procedures as written, training typically does little to help pilot understand the reasons they are vulnerable to errors in executing checklist and monitoring procedures. This is a crucial oversight, because individuals are better motivated and better prepared to deal with error-prone situations if they understand the nature of this vulnerability and the circumstances in which it occurs.

Suggestion: Pilots should be trained on their inherent vulnerability to checklist and monitoring errors, and on procedural measures and practical techniques to counter it. Our report could be used to generate a module within CRM training. The module should explain that many errors such as "looking without seeing" are inadvertent, and that pilots are often completely unaware that their performance has eroded. Instructors should explain that the slow, deliberate approach to executing checklists goes against the natural grain, which is for highly practiced actions to become fast, fluid, and automatic, with little if any conscious oversight. Performing a checklist rapidly or from memory is not the mark of proficiency but of misjudgment. Thus the slow, deliberate approach requires practice and vigilance to become habit in line operations. The few extra seconds required to perform a monitoring task or a checklist deliberately are well worth the slight time cost.

Instructors should facilitate a frank discussion of operational pressures that work against deliberately paced execution of procedures, in particular, pressure for on-time completion of flights and the distracting effects of interruptions and concurrent task demands. Crews can then discuss how best to deal with these pressures without compromising safety.

CRM classes often include a section on workload management, but this section typically focuses on managing overload by prioritization and distribution of tasks. This material could be expanded to address timing of initiation of tasks such as checklists. Also, pilots are so accustomed to juggling several tasks concurrently they may not recognize the several ways in which multi-tasking increases error rates. The book by Loukopoulos et al. (2009) provides material that can be incorporated into training modules to help pilots understand and counter this vulnerability. For example, individuals find it difficult to believe they could forget to perform simple but crucial tasks, executed so often as to become habit—such as setting flaps to take-off position and checking this action by performing a checklist item.[11] Thinking themselves highly unlikely to omit such a task, pilots may underestimate vulnerability and not recognize situations (e.g., interruptions, distractions, and being forced to perform a procedure out of its normal sequence) that increase vulnerability; consequently they may be less motivated to develop the habit of deliberate pacing. Dismukes (2010) and Loukopoulos et al.

---

[11] These are examples of *prospective memory* errors. Prospective memory refers to needing to remember—and sometimes forgetting—to perform an intended action at a later time.

(2009) suggest additional practical techniques to counter this vulnerability; for example, in critical situations some tasks should be suspended to allow undivided attention to the critical task, such as crossing an active runway or taxiway. When interrupted or deferring a task, pilots should pause a moment to identify where and when they will return to the task. Creating reminder cues can be very helpful; some pilots clip their tie to the yoke to remind them to periodically check fuel transfer during re-balancing. As we have discussed, other pilots turn on their taxi lights after receiving clearance to land to help themselves remember later if they have been cleared. However, although we observed several pilots use this taxi light technique, only one actually checked the light at a specific point in the approach. Without the habit of checking the taxi light switch, its effectiveness as a cue is reduced.

During CRM training, presentation of accidents in which highly experienced pilots inadvertently failed to execute a habitual task may help put the issue in proper perspective. But in order to effect lasting change in pilot performance on the line, all of the academic training discussed here must be reinforced in initial and recurrent simulator training and in line checks.

Suggestion: Reinforce the responsibility of monitoring pilots to challenge deviations. Even when pilots monitor appropriately, challenging deviations by the pilot flying often does not occur, for reasons previously discussed. Our findings, like those of previous researchers (Orasanu et al., 1999; Fischer & Orasanu, 2000) reveal that first officers are less likely to challenge a captain flying than vice versa; thus the airlines need ways to support challenging when appropriate—simply telling first officers to challenge is not sufficient to counter their hesitation. Both initial and recurrent training should address the issue realistically, which requires frank discussion of the reasons challenging is sometimes difficult. Pilots—especially first officers—must balance the need to challenge with maintaining a positive cockpit environment. An outstanding technique used by some captains during the initial briefing to a first officer goes something like: "I expect I will make errors on this flight—it is your job to catch them and point them out". Not only does this approach give the first officer permission to speak up, it establishes an atmosphere in which either pilot can challenge the other without causing him or her to lose face, and it establishes the standard that monitoring is an essential cockpit procedure.

Although captains more frequently challenged the flying pilot's deviations than did first officers, out results show that they too failed to catch most deviations; thus they too need more effective training. Regardless of their crew position, all pilots must be aware of the importance of remaining engaged in the monitoring task, even when operations are so routine that monitoring seems unnecessary, and even when workload is so high that it is tempting to abandon monitoring in favor of other task demands.

Suggestion: Develop techniques to provide detailed feedback to pilots on checklist and monitoring performance. Another reason pilots and other human operators do not recognize erosion of their checklist and monitoring procedures is that they rarely receive feedback when they make an error. As we noted throughout our observations, because the aviation system has many safeguards, pilot errors rarely result in consequences that bring the errors directly to the pilot's attention. In this respect the system operates open loop, yet to acquire and maintain good habits human operators require closed-loop feedback about their performance. This is not an easy problem to solve—it is an ironic consequence of an aviation system that normally operates at high levels of safety. However, there are ways feedback can be provided. Facilitated debriefings after simulation training can encourage crew members to give each other feedback—conducting these debriefings in training

events may help develop a company culture in which constructive feedback from fellow pilots is normalized, making crews more likely to debrief their operational flights on their own. Also, flight simulators can be programmed to present unexpected faults that should be caught by careful checklist use or monitoring. During non-jeopardy recurrent training, instructors can pull one pilot of a crew aside and ask him or her to deliberately make an error, which the other pilot should catch. Research could develop new direct methods of providing feedback during training—for example, eye-tracking devices that would record how long pilots' gaze fixates on items being checked. (Fixations of less than 300msec are not sufficient to process information adequately.) More broadly, research is needed on techniques and devices to help pilots maintain monitoring reliably, especially when they are not directly and actively controlling the system being monitored.

Suggestion: Place greater emphasis on checklist use and monitoring in air carrier flight standards (line checking) programs. Individuals are quite good at seeing through their organization's rhetoric about how the individual is supposed to perform. Almost all organizations maintain that "Safety is our highest priority", but the implicit messages organizations give suggest that in reality other objectives often have equal or even greater priority. One crucial feedback loop pilots receive is from periodic check flights in which a check airman flies with the pilot to evaluate performance. We know of no research on the relative emphasis check pilots give to diverse aspects of pilot's performance; however our observations and LOSA observations typically identify more problems with checklist use and monitoring than do check flights. We know of two airlines that have replaced the traditional line check with a "Line check safety audit", which draws upon the LOSA concept and which emphasizes evaluation of monitoring and checklist use. In order to close the feedback loop and make the line check valuable beyond just evaluation of the observed crew's qualification to fly, the check airman's debriefing of the crew should address error-trapping, checklist use, and monitoring. Check airmen may need training in the most effective ways to provide this feedback.

Suggestion: Develop formal mentoring programs for new first officers. "Schoolhouse" training and initial operating experience (IOE) are essential, but are only the starting point for first officers to gain skills and effective habits. Simulation training does not capture the full range of operational complexity, especially the concurrent task demands that work against effective checklist use and monitoring. If a first officer is lucky, she or he will encounter captains who go out of their way to pass along their experience and insight for dealing with these situations, but this is a haphazard process. Formal mentoring programs might tap captains' expertise more systematically and provide more standardized guidance. Such programs would of course address all aspects of flying, but checklist use and monitoring/challenging might benefit especially. Pilots get immediate feedback from the aircraft if their handling skills are lacking (e.g., a bounced landing), but aircraft performance does not give direct feedback on pilots' shortcomings in checklist use, monitoring, or challenging.

6.3 System Design

Cockpit features, some already available, can assist checklist and monitoring performance. Something as simple as a mechanical device used by one major airline and some military transports for Before Takeoff and Before Landing checklists, can reduce vulnerability to losing one's place in the checklist and omitting items. The device displays the checklist items, and, as each item is performed, the pilots throw a toggle switch that turns off the light next to the item; thus a quick glance at the device tells the crew whether all items have been completed. Electronic checklists, used in the current generation of airliners, are a substantial advance (Boorman, 2001). (Unfortunately, only one of the flights we observed was on an aircraft with an electronic checklist,

so we could not compare deviation rates.) These come in two types: integrated electronic checklists, which sense the status of some (though not all) of the items and stand-alone electronic checklists, which do not sense status of items. With the integrated checklists, after pilots complete the flow and start the checklist, they can skip over the items already set (displayed in green in one manufacturer's version), going directly to any items yet to be completed (displayed in white). This reduces the number of checklist items, thus reducing opportunities for missing an item, and it helps pilots keep track of where they are in the checklist. However, pilots should guard against becoming so reliant on the electronic checklist that they become less rigorous about performing the flow before the checklist (if the airline's SOPs specify a flow-then-check procedure).

The next generation of electronic normal checklists should also reduce vulnerability to omission of entire checklists. For several, but not yet all checklists, a caution message with aural tone and a master caution light will alert the crew that a checklist has not been completed before moving to the next phase of flight. These electronic checklists also automatically insert an item deferred from one checklist into a later checklist, reducing vulnerability to forgetting interrupted or deferred items (a function already implemented on the current electronic checklists for non-normal procedures). Electronic displays already remind pilots when they have forgotten some procedural items; for example, turning the altimeter display amber if the crew forgets to transition to or from QNE. Further advances will likely occur when it is possible to sense the status of more flow/checklist items and as artificial intelligence provides intelligent agents for the cockpit. However, as cockpit automation becomes ever more capable and comprehensive, airlines and pilots will have to be even more careful to avoid over-reliance on automation and deterioration of pilots' primary skills. Also pilots must retain good checklist habits for occasions when they have to go back to paper checklists because electronic checklists are deferred for maintenance.

Cockpit automation comes with many benefits, but it can also introduce new problems (Billings, 1997; Sarter & Woods, 1994), such as automation mode confusion and automation complacency. In particular, pilots often fail to monitor mode status and mode changes displayed with alphanumerics on the primary flight display (Sarter et al., 2007). Several decades into the era of flight management systems, the logic of some automation modes, including vertical navigation and its associated automatic mode changes on many aircraft types, is complex, situationally dependent in a way that challenges even experienced pilots, and poorly annunciated. Research is needed to develop mode operations that are clear to pilots and ways of displaying mode status that better engage the attention of pilots, especially when the system changes modes without pilot command. More broadly, even though automation has enhanced situation awareness in some ways, such as navigation displays, it has undercut situation awareness by moving pilots from direct, continuous control of the aircraft to managing and monitoring systems, a role for which humans are poorly suited. Also, the very reliability of automation makes it difficult for pilots to force themselves to "stay in the loop". Research is needed to develop ways to help pilots stay in the loop on system status, aircraft configuration, flight path, and energy state. These new designs must be intuitive and elicit attention as needed, but minimize effortful processing that competes with the many other attentional demands of managing the flight.

7. CONCLUSION

Although this study focused on deviations from prescribed procedures, these deviations must be understood in context. The vast majority of the actions of the observed crews were correct and effective and demonstrated required skills. Given the large numbers of opportunities for deviation, the deviation rates were probably well below one percent. We observed many examples of

exemplary performance and of effective techniques used to manage the challenges of cockpit operations. For example, several captains and first officers, by thinking ahead, identified possible consequences of existing or potential threats and acted preemptively to prevent those consequences. We also observed instances of very effective monitoring in which pilots caught a deviation made by the other pilot through overall awareness and scanning that was not part of an established SOP, flow, or checklist.

Even though modern airlines operate at extremely high levels of safety, the very fact that the level of safety is so high makes it difficult to detect when safety begins to erode. The tendency of any highly organized system is to become less well organized (using a metaphor from physics, entropy increases); thus, constant effort is required to maintain safety. The industry is under extreme pressure to cut costs, and the consequences of changes to training and procedures do not always show up immediately.

Our findings point to things that can be improved. In particular, trapping of errors and other deviations appears not to be operating at the level generally assumed. Most people in the airline industry now recognize that it is impossible to eliminate all human error, and that it is necessary to help pilots detect and manage errors before they become consequential. Threat and error management (TEM) programs are now fairly common, and many airlines address the need for cockpit monitoring. Yet these well-intentioned efforts appear to be falling short. We have suggested countermeasures that could provide a path to improvement; however, one limitation of our study approach is that it was by its nature phenomenological. We could observe and document crew performance and draw upon existing scientific knowledge to conjecture about the situational, cognitive, and organizational factors making pilots vulnerable to both inadvertent and intentional deviations from prescribed procedures. However, we did not have the opportunity to discuss these deviations with the crews to gain their perceptions. Other types of research are needed to extend our findings. For example, Mumaw, Roth, Vicente, and Burns (2000) supplemented observation of monitoring by nuclear power plant operators with extensive interviews. Experimental research is also needed to evaluate our conjectures about the factors underlying vulnerability to deviations and errors and to test the effectiveness of proposed countermeasures. Close collaboration between researchers and the aviation community is required for practical application of these countermeasures.

8. REFERENCES

AAIASB (2006). Aircraft accident report: Helios Airways Flight HCY522, Boeing 737-31S at Grammatiko, Hellas on 14 August 2005. Air Accident Investigation & Aviation Safety Board. The Hellenic Republic.

Besco, R. O. (1995). Releasing the hook on the copilot's catch 22 (crew members decision making). In Proceedings of the 39th Annual Meeting of the Human Factors and Ergonomics Society, pp. 20-24, San Diego: CA.

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

Boorman, D. (2001). Safety benefits of electronic checklists: An analysis of commercial transport accidents. In Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Degani, A. & Wiener, E. L. (1993). Cockpit checklists: Concepts, design, and use. Human Factors 35(2), 345-359.

Degani, A. & Wiener, E. L. (1994). Philosophy, policies, procedures, and practices: The four 'P's of flight deck operations. In N. Johnston, N. McDonald, & R. Fuller (Eds.), Aviation psychology in practice. Hants, England: Avebury Technical.

Dismukes, R. K. (2010). Remembrance of things future: Prospective memory in laboratory, everyday, and workplace settings. In D. Harris (Ed.), Reviews of human factors and ergonomics, Vol. 6. Santa Monica, CA: Human Factors and Ergonomics Society.

Dismukes, R. K., Berman, B. A., & Loukopoulos, L. D. (2007). The limits of expertise: Rethinking crew error and the causes of airline accidents. Burlington, VT: Ashgate.

FAA (2003). Standard operating procedures for flight deck crewmembers. AC120-71A, 27 February 2003. Federal Aviation Administration.

Fischer, U. & Orasanu, J. (2000). Error-challenging strategies: Their role in preventing and correcting errors. In Proceedings of the IEA2000/HFES 2000, pp. 30-33, San Diego: CA.

Flight Safety Foundation (2010). ALAR Toolkit. Retrieved 9 February 2010 from http://flightsafety.org/current-safety-initiatives/approach-and-landing-accident-reduction-alar/alar-tool-kit-and-resources.

Foushee, H. C., Lauber, J. K. Baetge, M. M. & Acomb, D. B. (1986). Crew factors in flight operations III: The operational significance of exposure to short-haul air transport operations. Houston, TX: National Aeronautics and Space Administration. NASA Technical Memorandum 88322.

Gross, R. L. (1995). Studies suggest methods for optimizing checklist design and crew performance. Flight Safety Digest, May 1995, 1-10 Flight Safety Foundation, Alexandria: VA.

International Civil Aviation Organization (ICAO). (1994). Safety analysis: Human factors and organizational issues in controlled flight into terrain (CFIT) accidents, 1984–1994. Montreal, Quebec, Canada.

Klinect, J. R., Wilhelm, J. A., & Helmreich, R. L. (1999) Threat and error management: Data from line operations safety audits. In R. S. Jensen (Ed.), Proceedings of the Tenth International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Lautmann, L. G. & Gallimore, P. L. (Oct 1987). Control of the crew-caused accident. Airline Pilot, pp. 10-14.

Li, G., Grabowski, J. G., Baker, S. P., & Rebok, G. W. (2006). Pilot error in air carrier accidents: Does age matter? Aviation Space & Environmental Medicine 27(7) 737-741.

Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2009). The multitasking myth: Handling complexity in real-world operations. Burlington, VT: Ashgate.

Mumaw, R. J. Roth, E. J., Vicente, K. J. & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. Human Factors 42(1), 36-55.

NTSB (1994). A review of flightcrew-involved, major accidents of U.S. air carriers, 1978 through 1990. Safety Study NTSB/SS-94/01. Washington, D.C.

Orasanu, J. McDonnell, L. K. & Davison, J. (1999). How do flight crews detect and prevent errors? Explanations for failures to correct errors. In R. Jensen (Ed.), Proceedings of the Tenth International Symposium on Aviation Psychology, pp. 350-355. Columbus, OH: Ohio State University.

Reason, J. (1990). Human Error. Cambridge, UK: Cambridge University.

Sarter, N. B. & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. The International Journal of Aviation Psychology, 4, 1-28.

Sarter, N. B. & Alexander, H. M. (2000). Error types and related error detection mechanisms in the aviation domain: An analysis of aviation safety reporting system incident reports. The International Journal of Aviation Psychology 10(2), 189-206.

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots monitoring strategies and performance on automated flights decks: An empirical study combining behavioral and eye-tracking data. Human Factors, 49, 347-357.

Shappell, S., Detwiler, C., Holcomb, K. Hackworth, C. Boquet, A., & Wiegmann, D. A. (2007). Human error and commercial aviation accidents: An analysis using the human factors analysis and classification system. Human Factors, 49, 227-242.

Sumwalt, R. L. (1999). Enhancing flight crew monitoring skills can increase flight safety. Flight Safety Digest, March 1999, pp. 1-8.

Sumwalt, R. L. III, Thomas, R. J. & Dismukes, R. K. (2002). Enhancing flightcrew monitoring skills can increase flight safety. In Proceedings of the 55th International Air Safety Seminar; Flight Safety Foundation (pp. 175-206), Dublin, Ireland, November 4-7. Retrieved 7 January 2008 from http://humanfactors.arc.nasa.gov/flightcognition/Publications/FSF_Monitoring_FINAL.pdf.

Sumwalt, R. L. III, Thomas, R. J. & Dismukes, R. K. (2003). The new last line of defense against aviation accidents. Aviation Week & Space Technology, 159(8), 66.

Thomas, M. J. W. (2004). Predictors of threat and error management: Identification of core nontechnical skills and implications for training systems design. The International Journal of Aviation Psychology, 14, 207-231.

Thomas, M. J. W. & Petrilli, R. M. (2006). Crew familiarity: Operational experience, non-technical performance, and error management. Aviation, Space, & Environmental Medicine, 77(1), 41-45.

Turner, J. W. & Huntley, M. S. (1991). The use and design of flightcrew checklists and manuals. Final Report AD-A237 206. U.S. Department of Transportation/Federal Aviation Administration.

Turner, T. P. (2001). Controlling pilot error: Checklists and compliance. McGraw-Hill: New York.

TABLE 1. NUMBER OF OBSERVED FLIGHTS BY COMPANY AND AIRCRAFT TYPE

Aircraft Type	*Company*			
	1	*2*	*3*	*Total*
A320	–	2	9	11
B737	29	–	–	29
B757	7	–	–	7
B767	–	–	2	2
B777	1	–	–	1
EMB 175/195	–	–	10	10
Total	37	2	21	60

Note: Dashes indicate aircraft type was not observed for that company.

	Checklists	Monitoring	Primary Procedures	Total
TABLE 2. DEVIATIONS PER FLIGHT: 3 MAJOR CATEGORIES				
Mean	3.2	6.5	5.2	15.0
Median	2.0	6.0	4.0	13.5
SD	2.9	3.7	4.9	8.2
Range	0–13	1–18	0–21	1–38

TABLE 3. DEVIATIONS IN EACH PHASE OF FLIGHT

Phase of Flight	Number of Deviations (Percent of Total)			
	Checklists	*Monitoring*	*Primary Procedures*	*Total*
Preflight	83 (9.2)	35 (3.9)	53 (5.9)	171 (19.0)
Taxi-out	20 (2.2)	19 (2.1)	39 (4.3)	78 (8.7)
Take off/ initial climb	0 (0)	13 (1.4)	10 (1.1)	23 (2.6)
Climb	8 (0.9)	164 (18.2)	33 (3.7)	205 (22.8)
Cruise	3 (0.3)	24 (2.7)	48 (5.3)	75 (8.3)
Descent	33 (3.7)	104 (11.6)	73 (8.1)	210 (23.4)
Approach	30 (3.3)	26 (2.9)	33 (3.7)	89 (9.9)
Landing	0 (0)	0 (0)	2 (0.2)	2 (0.2)
Taxi-in	6 (0.7)	2 (0.2)	20 (2.2)	28 (3.1)
Shut-down/ parking	11 (1.2)	4 (0.4)	3 (0.3)	18 (2.0)
Total	194 (21.6)	391 (43.5)	314 (34.9)	899 (100)

TABLE 4. COMPARISON OF NUMBER OF CHECKLIST ITEMS WITH NUMBER OF CHECKLIST DEVIATIONS

The number of checklists for one particular aircraft type at one airline is listed for each phase of flight. The number of challenge items and response items are the total from all checklists in a given phase of flight. The sum of challenge and response items is compared with the total number of checklist deviations observed in the study.

Phase of Flight	Number of Checklists	Number of Challenge Items	Number of Response Items	Sum of Challenge + Response Items	Checklist Deviations
Pre-taxi	3	33	47	80	83
Taxi-out	1	9	13	22	20
Take off/ initial climb	0	0	0	0	0
Climb	1	3	6	9	8
Cruise	0	0	0	0	3
Descent	1	7	12	19	33
Approach	2	8	14	22	30
Landing	0	0	0	0	0
Taxi-in	1	10	12	22	6
Parking	1	10	13	23	11

TABLE 5. TYPES OF CHECKLIST DEVIATION

Type of Deviation	Number of Deviations Observed	Percent of Checklist Deviations
Flow-check performed as read-do	48	25%
Responded without looking	43	22%
Item omitted, performed incompletely, or performed incorrectly	42	22%
Checklist initiated at poor time	31	16%
Checklist performed from memory	17	9%
Checklist not initiated	13	7%
Total	194	101% *

*The total is greater than 100 because of rounding to the nearest whole number.

TABLE 6. TYPES OF MONITORING DEVIATION		
Type of Deviation	*Number of Deviations Observed*	*Percent of Monitoring Deviations*
Callout late or omitted	211	54
Verification omitted	113	29
Not monitoring aircraft state or position	67	17
Total	391	100

TABLE 7. PRIMARY PROCEDURE DEVIATIONS			
Type of Deviation	*Deviation Sub-Type*	*Number of Deviations Observed*	*Percent of Primary Procedure Deviations*
Coordination	Crew-crew	56	18
	Crew-ATC	33	10
	Crew-ground personnel	8	3
	Crew-flight attendants	6	2
Configuration	Systems	62	20
	Aircraft	4	1
Planning or execution	Contingency	57	18
	Profile	7	2
Automation operation	Flight management system	40	13
	Mode control panel	18	6
	Head-down with automation too long	2	< 1
Path/airspeed control	Lateral	7	2
	Vertical	3	1
	Airspeed	1	< 1
Approach stabilization		10	3

	Captain as Flying Pilot	*Captain as Monitoring Pilot*	*First Officer as Flying Pilot*	*First Officer as Monitoring Pilot*
Mean	4.7	5.3	4.2	4.4
Median	3.0	4.5	3.5	4.0
SD	3.8	3.8	3.0	2.6
Range	0–15	1–17	1–12	0–10

TABLE 8. TOTAL DEVIATIONS PER FLIGHT BETWEEN
TAKEOFF AND LANDING AS A FUNCTION OF PILOT ROLE

TABLE 9. NUMBER OF DEVIATIONS PER FLIGHT FOR CREWS ON
FIRST FLIGHT TOGETHER VERSUS CREWS NOT ON
FIRST FLIGHT TOGETHER

	Checklists	*Monitoring*	*Primary procedures*
First flight together	4.3	8.0	10.1[a]
Not first flight together	3.2	6.4	4.7
First day together	3.4	7.4	7.8*
Not first day together	3.3	6.2	4.4

* Significantly higher than not first time together (two-tailed t-test).

TABLE 10. PERSON TRAPPING DEVIATIONS

Deviation Trapped by	Number	Percent
No one	738	82.1
Captain	64	7.1
First Officer*	65	7.2
Flight Attendant	2	.2
Observer	11	1.2
ATC	17	1.9
Aircraft system	2	.2
Total	899	100.0

* First officers were the monitoring pilot on 37 of the 60 flights and thus had more opportunities to trap the flying pilot's errors. First officers acting as the monitoring pilot trapped only 12.1% of the flying pilots' errors, whereas captains acting as the monitoring pilot trapped 27.9%. This difference between captains' and first officers' error trapping performance as the monitoring pilot was significant (two-tailed t-test).

TABLE 11. DEVIATION TRAPPING BY DEVIATION TYPE

Deviation Category	Specific Type of Deviation	Number of Instances	Number Trapped	Percent Trapped
Monitoring	Callout late or omitted	211	12	5.7
	Not monitoring aircraft state or position	67	9	13.4
	Verification omitted	113	1	0.9
	Total	391	22	5.6
Checklists	Flow-check as read-do	48	1	2.1
	Responded without looking	43	7	16.3
	Item omitted/incomplete/incorrect	42	6	14.3
	Poor timing	31	4	12.9
	Performed from memory	17	0	0
	Not initiated	13	10	76.9
	Total	194	28	14.4
Primary Procedures	Systems configuration	62	32	51.6
	Contingency planning/execution	57	3	5.3
	Crew–crew coordination	56	5	8.9
	Automation–FMS	40	16	40.0
	Crew–ATC coordination	33	25	75.6
	Automation–MCP	18	14	77.8
	Conducting unstabilized approach	10	0	0
	Crew–ground personnel coordination	8	0	0
	Profile planning/execution	7	4	57.1
	Lateral path control	7	3	42.9
	Crew–Flight attendant coordination	6	3	50.0
	Aircraft configuration	4	3	75.0
	Vertical path control	3	2	66.7
	Automation–head-down	2	0	0
	Airspeed control	1	1	100.0
	Total	314	111	35.4
Grand Total		899	161	17.9

TABLE 12. TYPES OF UNDESIRED AIRCRAFT STATES (UAS)
OBSERVED IN 31 SAMPLED FLIGHTS

Undesired State	Number of Instances	Percent of Total
Systems misconfigured	10	22
Airspeed incorrect	7	16
Unstabilized approach	5	11
Fuel below reserve	3	7
Vertical path deviation	3	7
Flight attendants not seated Takeoff or landing Turbulence	2 2	4 4
High and fast on approach	2	4
Hot brake not addressed	2	4
Landing from unstabilized approach	2	4
Navaid not identified and flight attendants not seated on approach	1	2
Aircraft controls misconfigured	1	2
Heading incorrect	1	2
Heading set incorrectly for takeoff	1	2
Lights off during climb	1	2
Excessive stopping distance	1	2
Terrain separation inadequate	1	2

TABLE 13. DEVIATIONS RESULTING IN UNDESIRED AIRCRAFT STATE IN 31 SAMPLED FLIGHTS

Deviation Category	Specific Type of Deviation	Number of Instances
Monitoring	Not monitoring aircraft state or position	5
	Verification omitted	3
	Callout late or omitted	2
Checklists	Item omitted/incomplete/incorrect	2
	Flow-check as read-do	1
	Responded without looking	1
	Timing	1
Primary procedure	Systems configuration	7
	Contingency planning/execution	5
	Unstabilized approach	4
	Automation–MCP	3
	Crew-ATC coordination	2
	Automation–FMS	2
	Lateral path control	2
	Poor profile planning/execution	2
	Crew-flight attendant coordination	1
	Aircraft configuration	1
	Total	44

1. REPORT DATE (DD-MM-YY) 07-23-10	2. REPORT TYPE Technical Memorandum	3. DATES COVERED (From – To)

4. TITLE AND SUBTITLE Checklists and Monitoring in the Cockpit: Why Crucial Defenses Sometimes Fail	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S) R. Key Dismukes and Ben Berman	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER WBS 031102.02.01.35.466A.10

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESSES(ES) NASA Ames Research Center Moffett Field, CA 94037	8. PERFORMING ORGANIZATION REPORT NUMBER TH-084

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001	10. SPONSORING/MONITOR'S ACRONYM(S) NASA
	11. SPONSORING/MONITORING REPORT NUMBER NASA/TM 2010-216396

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified—Unlimited
Subject Category: 03
Availability: NASA CASI (301) 621-0390

14. ABSTRACT

Checklists and monitoring are two essential defenses against equipment failures and pilot errors. Problems with checklist use and pilots' failures to monitor adequately have a long history in aviation accidents. This study was conducted to explore why checklists and monitoring sometimes fail to catch errors and equipment malfunctions as intended. Flight crew procedures were observed from the cockpit jumpseat during normal airline operations in order to: 1) collect data on monitoring and checklist use in cockpit operations in typical flight conditions; 2) provide a plausible cognitive account of why deviations from formal checklist and monitoring procedures sometimes occur; 3) lay a foundation for identifying ways to reduce vulnerability to inadvertent checklist and monitoring errors; 4) compare checklist and monitoring execution in normal flights with performance issues uncovered in accident investigations; and 5) suggest ways to improve the effectiveness of checklists and monitoring. Cognitive explanations for deviations from prescribed procedures are provided, along with suggestions for countermeasures for vulnerability to error.

15. SUBJECT TERMS

checklist use; monitoring; airline operations

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON STI Help Desk at email: help@sti.nasa.gov
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU	62	19b. TELEPHONE NUMBER (Include area code) STI Help Desk at: (301) 621-0390