QUANTIFYING PILOT CONTRIBUTION TO FLIGHT SAFETY FOR NORMAL AND NON-NORMAL AIRLINE OPERATIONS

Timothy J. Etherington, Rockwell Collins, Hampton, Virginia
Lynda J. Kramer, Randall E. Bailey, Kellie D. Kennedy, Chad L. Stephens
NASA Langley Research Center, Hampton, Virginia

Abstract

Accident statistics cite the flight crew as a causal factor in over 60% of accidents involving transport category airplanes. Yet, a well-trained and well-qualified pilot is acknowledged as the critical center point of aircraft systems safety and an integral safety component of the entire commercial aviation system. No data currently exists that quantifies the contribution of the flight crew in this role. Neither does data exist for how often the flight crew handles non-normal procedures or system failures on a daily basis in the National Airspace System.

A pilot-in-the-loop high fidelity motion simulation study was conducted by the NASA Langley Research Center in partnership with the Federal Aviation Administration (FAA) to evaluate the pilot’s contribution to flight safety during normal flight and in response to aircraft system failures. Eighteen crews flew various normal and non-normal procedures over a two-day period and their actions were recorded in response to failures. To quantify the human’s contribution, crew complement was used as the experiment independent variable in a between-subjects design. Pilot actions and performance when one of the flight crew was impaired were also recorded for comparison against the nominal two-crew operations. This paper details a portion of the results of this study.

Keywords – flight crew error, aviation safety, increasingly autonomous systems, automation, automation surprise, crew complement, augmented crew, single pilot operations.

Introduction

Accident statistics and many analytical efforts cite that flight crew error is a primary contributor in accidents involving transport category airplanes [1]. However, others suggest “a well-trained and well-qualified pilot is acknowledged as the critical center point of aircraft systems safety and an integral safety component of the entire commercial aviation system” [2]. The latter statement is generally acknowledged but cannot be verified because little or no quantitative data exists on how or how many accidents/incidents are averted by crew actions. In fact, anecdotal evidence suggests crews handle routine failures on a daily basis and airline’s Aviation Safety Action Program (ASAP) data [3, 4] supports this to be true although the data are not publicly released. In the absence of hard data, the pilot’s contribution to and the methods by which the human operator improves safety-of-flight is difficult to positively define. Ways to improve a human’s ability to provide this positive contribution are similarly ill-defined.

NASA is conducting research [5] to address the needs of future air transportation and airspace operations. Under its strategic plan, NASA is conducting research, development, test, and evaluation to create “assured autonomy for aviation transformation.” The Safe Autonomous Systems Operations (SASO) Project [6] is charted to identify and develop the possible autonomous capabilities enabling operational concepts and complexity that sustain significant growth in passengers and cargo beyond the Next Generation Air Transportation System (NextGen). The project’s goal is to seek ways to safely integrate within the National Airspace System (NAS) the highest level of autonomy that is justifiable, and develop methods for assurance. The work is premised that future airspace capacity and safety will be restricted by human performance constraints or limitations.

The SASO research portfolio aligns with the Assured Autonomy for Aviation Transformation strategic thrust area and the guidance provided by the National Research Council [7] which called for research in Increasingly Autonomous Systems (IAS). IAS are envisioned as intelligent machines -
hardware and software systems - seamlessly integrated with humans whereby task performance of the combined system is significantly greater than the individual components. IAS utilize machine learning concepts to perform functions without explicitly being programmed. These systems have the ability to modify their behavior in response to their external environment and conditions. Nevertheless, these IAS are also designed using human-autonomy and automation interaction teaming principals, without the active and latent failings prevalent in many of today’s automated cockpits.

Since in the design of IAS, “neither the general public nor the aviation community would tolerate any decrement in safety as a result of changes in the NAS associated with introduction of advanced Increasingly Autonomous Systems” [7], it is incumbent that the factors necessary to the safe operation of vehicles in the NAS be understood and that methods by which these safety factors are obtained are not adversely affected. In fact, IAS, if properly designed, can replicate and in fact should, enhance safety and reliability [7] (e.g., provide the ability to adapt to changing patterns and preferences; remain vigilant at all times; tailor automation actions to specific circumstances/add flexibility; increase situation awareness by context-sensitive information; monitor human actions and alert or intervene to prevent errors; improve automations’ ability to react quickly to avoid critical situations).

NASA Langley Research Center conducted a study to provide data on the contribution of the flight crew to aviation safety. This study was designed to provide data on how flight crews handled normal and non-normal conditions during typical revenue-like flight operations. These data are fundamental to and critical for the design and development of future IAS that can better support the human in the cockpit. Different crew complements were studied to gain understanding of the safety afforded by having two crewmembers on the flight deck. Normal two-crew operations were contrasted and compared to conditions were the second crew member was unavailable when the non-normal condition occurred but became re-engaged after returning to the flight deck and another case where only a single pilot was on the flight deck simulating an incapacitated pilot. This paper details preliminary results and analysis of one of six non-normal events tested.

Method

Experiment Design

The test objectives of the experiment were as follows:

1. Establish “baseline” levels of performance and safety with nominal two-crew configuration as well as collecting data to assess the performance and safety decrements in reduced crew and single pilot crew complements for present-day flight deck design and certification; and,

2. Identify technology requirements from these data for increasingly autonomous systems that might assist future two-crew operations and eventually, enable reduced crew or ultimately, single pilot operations.

To assess human performance and safety the experiment contrasted two-crew operations to conditions when one of the pilots was absent from the flight deck. If the condition was temporary, we designated it reduced crew operations (RCO) - one of the pilots was resting in place on the flight deck. If the condition was permanent, like pilot incapacitation, we designated that single pilot operations (SPO). These terms are used throughout this document.

Participants

Thirty-six pilots (18 crews), representing five airlines, participated in this experiment. Each pilot held an Airline Transport Pilot rating and was current in the 737-800 aircraft as either Captain (CP) or First Officer (FO). Crews were paired from the same employer to minimize inter-crew conflicts in Standard Operating Procedures (SOPs) and Crew Resource Management (CRM) procedures or training. All participants were male. The Captains’ average age was 55.1 years with an average of 3,508 flight hours flying as 737-800 CP, 21.3 years of commercial flying, and 15.6 years (for 12 of the 18 pilots) of military flying. The First Officers’ average age was 47.1 years with an average of 2,073 flight hours flying as 737-800 FO, 10.1 years of commercial flying, and 19 years (for 8 of the 18 pilots) of military flying. Seven of the CPs and three
of the FOs were Instructor Pilots (IPs) for their respective air carriers. Two CPs were trained in the US Navy as test pilots and one of these CPs was also flying as a test pilot for his air carrier.

Subjects were instructed to bring their company’s electronic charts and 737-800 checklists with them. Familiar charts and checklists further reduce conflicts in SOPs and CRM procedures or training.

Pilot flying (PF) and pilot monitoring (PM) roles were assigned based on the experimental matrix.

**Apparatus**

The research was conducted using the B-737-800 simulator operated by the FAA AFS-440 at Oklahoma City, OK (Figure 1). The simulator is a replica of a B737-800 aircraft, principally following American Airlines equipage. It is Level D certified so it can be used for both initial and recurrent training. As such, it includes the assemblage of equipment and computer programs necessary to replicate aircraft operations in ground and flight conditions, a visual system providing an out-of-the-cockpit view, a system that provides cues to mimic flight and ground operations, and has the full range of capabilities of the systems installed in the device sufficient for full initial and recurrent training.

![Figure 1. FAA OKC B737-800 Simulator](image)

The simulated weather en-route contained significant areas of convective activity along the Rocky Mountain Front Range and strong Northerly winds which required a north departure out of KDEN before a circuitous route to the west and then south to KABQ. This same planned route of flight was used for the entire two days.

A live controller and a pseudo-pilot were tied into the simulation radio in real-time to simulate Air Traffic Control (ATC) and some proximate traffic to promote realism and maintain realistic pilot workload levels. The subject pilots were able to visually acquire the traffic out the window (weather conditions permitting) and also on Traffic Collision and Avoidance System (TCAS). The pseudo-pilot also served as a dispatcher in the Airline Operations Center and provided communications as necessary and appropriate with own-ship during the scenarios.

**Training**

The crews were current and qualified B737-800 pilots. The simulator was Level D certified, therefore, no additional training was conducted.

There were no significant cockpit and aircraft differences between the airlines current operation and the FAA B737-800 simulator. One possible exception was the difference in the Electronic Flight Bag (EFB) mounting installation that sometimes differ amongst airlines. The simulator was head-up display (HUD)-equipped. Several operators, but not all, have installed HUDs and their use is sometimes prevalent. To ensure commonality, the crew was instructed that the HUD would *not* be used and it was stowed.
The crews were briefed on the purpose of the experiment and received the dispatch paperwork. The crews were instructed to use their normal company SOPs and checklist procedures for the entire test, including any company dispatch calls and cabin crew communications. The aircraft Auxiliary Power Unit (APU) was not in service for the flight; this was noted in the dispatch paperwork. The APU is not on the Minimum Equipment List and crews were authorized to dispatch with the APU inoperative.

After the crews had reviewed the paperwork and discussed amongst themselves the flight and flight conduct, the crew manned the aircraft, did a quick familiarization check and reviewed the simulator safety briefing. Known simulator-isms and aircraft differences were identified and discussed with the crew. The aircraft initial condition was in the hold-short of Runway 35L at KDEN with the engines running, parking brake set. (Push-back from the gate was not used as the starting point to save some simulation time without loss of fidelity.) The Flight Management System (FMS) was already loaded with the planned flight routing; the crews were asked to double check the entries. After review and confirmation of the cockpit switches and cockpit setup and completing their normal checklists, the crew called KDEN tower for departure.

Following clearance from ATC, the crew flew an entire “nominal” flight from KDEN to KABQ following the planned route of flight. The nominal flight served as a baseline for ‘normal’ airline operations (i.e., nominal data) and also helped to garner familiarity for the two-person crew interaction. This nominal flight was 1+25 hrs of flight time.

**Independent Variables**

The independent variables were crew complement and scenario. Three crew conditions were used: two-crew; RCO; and SPO and six non-normal scenarios.

**Crew Complement Configurations**

For the two-crew condition, the CP and FO were seated in the left and right seats, respectively.

For the RCO condition, the CP and FO were seated in the left and right seats, respectively, as usual. Whichever pilot was to serve as the resting pilot, remained seated in their seat, but was wearing a visual/audio restriction device (VARD) to be isolated from the sights and sounds of the flight. The device was sufficient that most pilots did not report hearing master caution, autopilot disconnect, or overspeed clacker alerts while resting.

For the SPO condition, the CP or FO were seated in the left seat. Although this is not the usual seat for the FO, this methodology was planned because the left seat has access to all normally required equipage (e.g., tiller, etc.) as would be expected for an aircraft designed for SPO. The pilot-not-flying was not in the simulator.

**Scenario Categories**

Six scenarios were designed, in pairs, for the following failure mode groupings:

<table>
<thead>
<tr>
<th>Category</th>
<th>Root Cause</th>
<th>Ancillary Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Unalerted System Malfunction</td>
<td>Autopilot available</td>
</tr>
<tr>
<td>B</td>
<td>Alerted System Malfunction</td>
<td>Autopilot available</td>
</tr>
<tr>
<td>C</td>
<td>Alerted System Malfunction</td>
<td>Autopilot not available</td>
</tr>
</tbody>
</table>

The categories differed whether the pilots could use the autopilot after the failure and whether the failures were immediately alerted automatically for the crew. These two effects were used to identify workload and automation issues (i.e., by availability of autopilot) and flight crew awareness and monitoring for normal / non-normal operations (i.e., alerting).

Two non-normal scenarios, described in the following, populated each category.

The experimental scenarios started at either the Top-of-Climb (TOC) or approximately 15 minutes prior to the Top-of-Descent point (i.e., TOD-15) positions along the nominal flight plan. The simulation was reset to these initial conditions and the crew/pilot were briefed on the starting position prior to going to “operate.” These two initial conditions were used for test expediency. The crews were briefed that the flight to this point was nominal.
The Category A and C non-normal conditions were triggered at a seemingly random point (10 to 20 minutes) after starting the simulation at the TOC point.

The Category B non-normal conditions were triggered at a seemingly random point (10 to 15 minutes) after starting the simulation at the TOD-15 point.

All flights were flown to completion (i.e., landing).

To meet the test objectives, the following test matrix was used (Table 1) with the number of populated cells (N) defined with the 18 crews:

<table>
<thead>
<tr>
<th>Table 1. Experimental Test Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenarios</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Nominal</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>Unreliable Airspeed</td>
</tr>
<tr>
<td>Engine Fuel Leak</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>Reservoir Hydraulic Leak</td>
</tr>
<tr>
<td>Generator Drive Shaft Failure</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>Loss of Both Generators</td>
</tr>
<tr>
<td>Rudder Trim Runaway</td>
</tr>
</tbody>
</table>

The experiment matrix (Table 1) shows that data were collected for all 18 crews flying a flight from KDEN to KABQ in a normal two-crew configuration, of which, the most important data was the arrival into KABQ. This same scenario was also flown by 9 crews, in an RCO configuration (6 with the CP resting and 3 with the FO resting) during the cruise portion of the flight. Finally, 18 pilots (9 FOs and 9 CPs) flew from the TOD-15 initial condition to landing and roll-out in a SPO crew condition (i.e., crews were not asked to fly an entire flight in the SPO configuration).

Some observations in the design:

1. The design is between subjects providing a direct comparison across each scenario (N=6) for the two-crew, CP resting, and SPO conditions;

2. The effect of scenario for a constant crew configuration is provided (N=6);

3. There is a direct comparison for two-crew and SPO for approach and landing (both CP and FO), N=18. This comparison is provided by the initial nominal data run (KDEN to KABQ) with two-crew, and by flying SPO with the CP and FO from the TOD-15 point on the same arrival. Familiarity of the first arrival was assumed to not significantly change performance.

4. Similarly, the RCO crew configuration was flown on the entire KDEN to KABQ nominal route in the RCO configuration (N=6 for the FO resting and N=3 for the CP resting). These data provide comparison to the two-crew configuration for the acceptability of concept of operation. Familiarity of the first arrival was assumed to not significantly change performance. The pilot who was resting was instructed to serve as the PF for the descent, approach, and landing phase of the flight.

5. The data for FO resting will be strictly observational since there is not enough power for statistical analysis. The observational data will, however, help inform subsequent testing associated with crew training and procedures.
6. No crew member saw a repeat of the same scenario with the exception of the nominal flight to and arrival into KABQ.

7. The run order was arranged so that the RCO crew conditions were flown one each in the A, B, and C scenarios. Similarly, the two-crew conditions for a crew were not flown in the same scenario category.

The initial two-crew data run (flown on Day 1 from KDEN to KABQ) provided nominal data and established the basis by which six additional experimental failure scenarios were evaluated.

In addition to the baseline, nominal flight, each crew flew six failure scenarios and an additional nominal flight over a two day period:

• Two of the failure scenarios in a Two-Crew configuration,

• Three of the failure scenarios in an RCO configuration,

• One of the failure scenarios in an SPO configuration with each pilot (both CP and FO) flying the scenario independently, resulting in two SPO runs for that particular failure scenario.

• Lastly, 9 crews flew the nominal arrival into KABQ (staring at the TOD-15 point and flying to roll-out and turn-off from the active KABQ runway) in an SPO configuration by each pilot, both CP and FO; the other 9 crews re-fl ew the baseline, nominal flight (from KDEN to KABQ) in an RCO configuration (6 with the CP resting and 3 with the FO resting), where the pilot rested during cruise from TOC to the TOD-15 points.

**Dependent Variables**

The dependent variables included both objective and subjective measures. The objective variables included complete performance parameters of the aircraft and selected switch positions and system operation. An event timer was provided for the researcher to record events such as failure recognition, checklist usage, and other observations that might require further data analysis using the objective measurements. A FitBit™ was used to measure heart rate [8] and eye tracking was used to quantify gaze.

Subjective measures were administered immediately after the completion of each run and at the end of all runs. These measures included the NASA Task Load Index (TLX) [9-11], Likert-type questions [12], semi-structured debriefing questions, and discussion. Participants were debriefed with a guided questionnaire and semi-structured discussions at the end of the final run. At the end of each scenario run, all crews provided verbal scores for four questionnaires; NASA TLX, Crew Resource Management (Figure 2) [13], Failure/Recovery Scale [14], and a Perceived Safety of Flight (Figure 3). During RCO runs only, two additional questions were asked regarding the reengagement of the resting pilot (Figure 4).

The communication between Captain/First Officer was appropriate and effective?

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Slightly Agree</th>
<th>Neutral</th>
<th>Slightly Disagree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

There was a shared awareness of the situation between the Captain/First Officer?

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Slightly Agree</th>
<th>Neutral</th>
<th>Slightly Disagree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

The decision-making process was appropriate and effective?

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Slightly Agree</th>
<th>Neutral</th>
<th>Slightly Disagree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

There was appropriate and effective allocation of tasks (i.e., prioritized, avoiding overload/underload, etc.)?

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Slightly Agree</th>
<th>Neutral</th>
<th>Slightly Disagree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

We performed as an effective team?

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Slightly Agree</th>
<th>Neutral</th>
<th>Slightly Disagree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

**Figure 2. CRM rating scale**
Perceived Safety of Flight Rating
Instructions: Please provide a numerical rating (1-7) for the following statement for the flight you just completed.

The perceived safety of flight during Dual-Crew/RCO/SPO operation is:

<table>
<thead>
<tr>
<th>Completely Acceptable</th>
<th>Neutral</th>
<th>Completely Unacceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Perceived Safety of Flight Rating

Resting Pilot Re-Engagement Ratings
Instructions: Please provide a numerical rating (1-7) for the following statement for the flight you just completed.

To safely and effectively respond to the non-normal scenario, the time that the resting pilot re-entered into the operation of the flight was:

<table>
<thead>
<tr>
<th>Too Early</th>
<th>Just Right</th>
<th>Too Late</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the non-normal scenario, the re-engagement of the resting pilot was:

<table>
<thead>
<tr>
<th>Completely Ineffective</th>
<th>Neutral</th>
<th>Completely Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Resting Pilot Re-Engagement Ratings

The NASA TLX captured a subjective rating (0 [Low] to 100 [High]) of perceived task load. There are six subscales of workload represented in the NASA TLX: mental demand, physical demand, temporal demand, performance, effort, and frustration level [9]. The overall and subscale score results of this measure were examined to investigate task load variation.

CRM was assessed with five questions using a seven-point Likert-type scale ranging from strongly agree (1) to strongly disagree (7). This metric was based on the CRM assessment methods [13] in AC120-51e, Section 14(a) and focused on the essential areas of CRM-related assessment: a) Communication; b) Decision-making; c) Team-building; d) Workload maintenance; and, e) Situation Awareness (Figure 2).

The effect of crew complement (two-crew, RCO, or SPO) on perceived safety of flight (Figure 3) was assessed on a seven point Likert-type scale ranging from completely acceptable (1) to completely unacceptable (7).

The Failure/Recovery Ratings metric [14] captured two ratings: Effect of Failure and Ability to Recover. These were assessed using a decision tree styled after the Cooper-Harper Pilot Rating Scale. The first rating – “The Effect of Failure” - ranged from “Minimal Excursions in Aircraft State” (rating of A) to “Catastrophic encounter with obstacles or structural failure” (rating of H). The second rating – “The Ability to Recover” - ranged from “Corrective control action not require” (rating of A) to “No possibility of averting catastrophe” (rating of H).

Following RCO runs only, subjects were asked two questions using a seven-point Likert-type scale to gauge the timing and effectiveness of the resting pilot re-engagement back into the simulation (Figure 4). Resting Pilot Timing ranged from too early (1) to too late (7). Resting Pilot Re-Engagement ranged from completely ineffective (1) to completely effective (7).

Results

The results shown here describe the major findings for only one of the failure categories, the rudder trim failure. The rudder trim failure is a Category C non-normal where the failure is alerted by an autopilot disconnect alarm once the trim runs to its maximum value and the autopilot can no longer compensate for this mis-trim. The autopilot is no longer available after the failure.

The rudder trim failure was un-annunciated initially, although several pilots noted the situation developing from wheel and rudder movement, until it resulted in a sudden loss of autopilot, and the aircraft immediately started to leave controlled flight. It appeared at first to be a roll problem. It is not a failure most pilots see in training and the only appropriate checklist is a generic checklist that covers all control problems. After the aircraft was under control, close examination provided cues of the real problem and when pilots looked at the rudder trim indicator, they correctly determined the actual problem. Because the checklist did not specifically cover the rudder trim failure, some crews decided that the checklist didn’t apply. The trim issue meant the autopilot was lost for the rest of the flight and the aircraft had to be hand-flown through approach and landing.

Rudder trim in the 737 is supplied through a spring in the rudder feel system. At full deflection, it provides around 40 pounds of force the pilot is required to counteract for the remainder of the flight.
For the rudder trim runaway failure event, 21 runs total were conducted with the following distribution of runs by crew configuration:

- 6 two-crew runs, with equal distribution of who (CP or FO) was designated as PF for the leg of the flight with the failure occurrence.
- 6 runs with CP as resting pilot during cruise operations. Descent, Approach, and Landing leg of flight was designated to the CP for these runs.
- 3 runs with FO as resting pilot during cruise operations. Descent, Approach, and Landing leg of flight was designated to the FO for these runs.
- 6 SPO runs, 3 flown by CPs and 3 flown by FOs.

One of the SPO-FO runs was excluded from the subjective measure analyses due to an unexpected action of the pilot. The FO asked dispatch for maintenance to read the Quick Reference Handbook (QRH) to him. The experiment team had not expected this request to be made and as such, the person acting as maintenance in this test had not been provided a QRH with the checklists. The pilot was mistakenly told to go to Standby Rudder by dispatch. When the pilot took this action, it made the aircraft unstable and the Principal Investigator (PI) ended the run. For the subsequent data collection runs, a QRH was accessible to any individual talking to the crew over the radios. Although this action created an unstable condition, the condition should have been recoverable and at least one other crew mistakenly did the same procedure (i.e., went to Standby Rudder) when diagnosing the problem and not using the proper checklist and they did not depart controlled flight.

Failure Handling

For the rudder trim runaway, the failure occurred approximately 20 minutes into the cruise phase of flight in central Colorado, heading south toward KABQ. A swept-wing aircraft is highly susceptible to slip-induced roll, which the rudder trim failure induced.

Upon failure, the pilot was required to counter the trim failure with rudder (~40 lbs) and counter-roll control. This level of force is difficult to maintain and especially difficult to maintain consistently.

Once the failure occurred, all 21 pilots/crews out, regardless of the crew complement, declared an emergency with ATC. Once the emergency was declared all pilots/crews requested some type of assistance from ATC and dispatch to help sort out the problem. The pilots/crew universally requested from ATC weather observations en-route and at possible divert fields as well as special handling (vectors for long straight-in approaches and block altitude clearances since aircraft maneuvering was compromised).

The test was staged to evaluate decision-making by the flight crew. The available divert airfields along the route were designed with relative merits and complications to tease out decision-making.

The diversion decision had to account for distance to fly with the failure, the weather at each airport and the potential for having control issues in bad weather and turbulence, and the time it took to troubleshoot the problem. Pilots are typically goal-oriented and to see if their decision was biased to their destination, the percent of diversions to other than the destination airport was calculated. Figure 5 shows the diversion airport by crew configuration and Figure 6 shows the percent of diversions to other than the destination of Albuquerque (KABQ).

The diversion decision reflects the overall pilot’s awareness of flight conditions and all available information. ATC and dispatch were available for weather information, and pilots were provided different information based on the quality of their request. They received a different answer if they asked for the closest airport than if they asked for where is the closest airport with visual flight conditions. The nearest airport to the failure was the departure airport (KDEN) however, there was significant convective activity along the route back to KDEN and the airport weather was overcast (500 ft ceiling with 1 mile visibility) and raining (i.e., wet runways). Weather at the destination (KABQ) was better (1000 ft ceiling and 3 miles visibility) with no convective weather near there. Grand Junction (KJGT) was 5,000 ft broken with ten mile visibility and dry runways and provided the best weather but it was not a normal airline destination. KJGT was also almost as far as KABQ after the Category C failure.
scenarios take place. The failures occurred over the mountain range and all closest airports were at a significant elevation of over 5000 feet MSL. For planning, this did not result in significant issues but did merit consideration if a flaps 15 landing was attempted where increased approach speed was required.

For the RCO configuration, the majority of the flights continued to the destination (KABQ). Only a third of the flights diverted to a different destination.

All of the single pilot flights where the FO was flying continued on the destination (KABQ) and none diverted. In the single pilot configuration with the CP flying, the majority of the flights diverted, split evenly between Denver, Grand Junction, and Albuquerque. The data reflects the captain’s experience level and decision-making ability.

Critical failures and difficult landings such as low visibility operations are often made by the CP. To see if this failure had any effect on the landing made by the FO or the CP, this data was analyzed.

There were three runs in the RCO-FO configuration for this failure event. It was the FO’s leg of the flight so he should have done the landing.

For this failure event, none of the FO’s made the landing. All landings were made by the CP.

For the RCO-FO nominal runs, all landings were made by the FO, as expected.

There were six runs in the RCO-CP configuration for this failure event. It was the CP’s leg of the flight so he should have done the landing.

For this failure event, 3 of the landings were conducted by the CP and 3 were done by the FO.

For the RCO-CP nominal runs, all landings were made by the CP, as expected.

There were 6 runs in the dual-crew configuration for this failure event.

For the 3 failure runs where it was the CP’s leg of the flight, all 3 landings being made by CP.

For the dual-crew nominal runs where it was the CP’s leg of flight, all landings were made by the CP.

Of the 3 failure runs where it was the FO’s leg of the flight, 2 of the landings were made by the FO and 1 by the CP.
For the dual-crew nominal runs where it was the FO’s leg of flight, all landings were made by the FO, as expected.

Normally, the Captain would perform landings in a suspect or damaged airplane. The nature of this failure provided some strange dynamics. First, since the rudder trim indicator supported what actually happened to the aircraft, many of the pilots felt they understood the failure and what affect it would have on aircraft performance in landing configuration. Some pilots even ignored the emergency landing guidance for a flaps 15 landing. Although these could be viewed as errors of commission, it does illustrate the complex nature of the scenario. Second, fatigue played a significant factor in who actually made the landing. Third, an interesting dynamic was if the failure happened to the FO, there was a tendency not to ask the CP for help, but if the CP experienced the failure, he was usually quick to ask the FO for help. Many crews switched back and forth and often the pilot that was flying closest to the destination airport was the pilot that did the landing. Some captains actually planned for the FO to fly to a particular point where they intended to take over for the landing. There was a tendency for the captain to do the landing with this failure but it was not universal.

A couple of pilots developed an interesting strategy. They found that if they jammed their foot under the rudder they could more easily counteract the force and they could hand-fly without difficulty.

**Workload**

As expected, mean workload as reflected in the CP and FO TLX ratings were greater during the rudder trim runaway failure event when compared to the nominal run TLX ratings for all crew configurations tested. The notched boxplots in Figure 7 indicate that the CP and FO TLX median workload ratings were significantly different (0.05 significance level) between the nominal runs and rudder trim runaway runs. The separation is almost 40 points out of a 100 scale.

In Figure 8, the workload components of the TLX workload measure for the PF are compared side-by-side between the nominal and the rudder trim runaway failure. These data are collapsed across crew complements. The boxplots of Figure 8 show the median ratings, with the 25th and 75th percentile spread in the data, as well as maximum and minimum values. The data primarily shows the workload differences between the nominal and rudder trim runaway failure. The Physical and Effort components of the TLX are most different, with spreads of 40 and 50 points, respectively, between the nominal and failure conditions. Although the failure case ratings were higher, the differences in Mental and Temporal demand between nominal and the failure case were about 20 points different in median rating. Performance was rated by the PF to be almost the same, although the large spread in the data for the rudder trim runaway indicates a large disparity in the self-assessments. Performance was only marginally decreased as all the fights but one ended in a successful landing and even though there were multiple excursions from nominal in roll, overall they handled a difficult situation as professional aviators.

**Figure 7: Overall TLX Ratings for CP and FO between Nominal and Rudder Trim Runaway**

**Safety-of-Flight**

Perceived level of safety was self-assessed using a Likert type scale from 1-7, where 1 is completely acceptable and 7 is completely unacceptable.
The notched boxplots in Figure 9 indicate that the PF’s median Perceived Safety of Flight ratings were significantly different (0.05 significance level) between the nominal runs and rudder trim runaway runs.

It is clear from an overall perceived level of safety for the PF compared to normal flight, this failure was difficult as expected. Also as expected, pilots viewed the safety of this failure as unacceptable during single pilot operations (mean rating=5.6). When handled single pilot, the perceived level of safety for a failure that requires the aircraft to be hand-flown with degraded handling qualities is completely unacceptable. This is tempered slightly by the experience level of the captains in this study. The mean rating during SPO for the FO was 6.5 and ranged from 6 to 7. The mean rating during SPO for the CP was 5 but ranged from 2 to 7.

Figure 8. TLX Components for CP between Nominal and Rudder Trim Runaway Collapsed across Crew Configuration.

Checklist Usage

Time to first correct checklist is a metric that should yield quick and proper troubleshooting of equipment problems. For the rudder trim failure, it was an alerted failure since the autopilot disconnected but there was no annunciation on the flight deck that had a direct entry in the QRH. This failure was additionally problematic in that there was no exact checklist procedure for a rudder trim failure. The proper and only checklist was Jammed and Restricted Flight Controls. Some crews elected not to do the procedure and this preliminary data reflects when they opened or made the decision that the checklist did not apply. Figure 10 shows the time to correct checklist usage based on crew configuration.

Figure 9. Perceived Safety of Flight Ratings for Pilot Flying between Nominal and Rudder Trim Runaway Collapsed across Crew Configuration.

Figure 10. Time to Correct Checklist

The data in Figure 10 clearly shows the rapid response with a two-crew configuration, a slightly
delayed response in the RCO case, and finally, significant latencies in the SPO case. In most cases for the RCO configuration, the PF waited until the resting pilot came back into the flight because, literally, he had his hands full. In the SPO case, getting access to the QRH, locating the proper items, and executing the checklist was extremely difficult for this failure.

Time to execute the correct checklist can be considered an indicator for safely handling failures. This metric was difficult to completely quantify with this failure due to the failure not being associated with an obvious checklist. Mean time for executing correct checklist for two crews was approximately three minutes, four minutes for the resting crew configuration and twenty minutes for the single pilot configuration. The time for the resting pilot to reengage on the flight deck was controlled by the researchers and was constant at two minutes. This two minutes is reflected in the slightly longer time for start of checklist usage in the resting pilot case. For the single pilot case, because the pilot was hand flying the aircraft and handling qualities were degraded, the time to conduct any meaningful checklist was significantly increased. If pilots reached for the QRH, the aircraft often started to go out of control. Even momentary lapses in concentration with this failure produced roll excursions. While single pilot, even the normal ATC communications interruptions significantly disrupted checklist flow. More than half the pilots in the single pilot configuration elected to do the descent and landing checklist from memory.

Flight Path Control

The number of times the crew exceeded a safe bank angle and the maximum bank angle exceeded illuminate how well the crew were able to recover from the initial failure as well as handling the failure for safe continued flight. Figure 11 shows the number of bank angle excursions beyond 30 degrees by crew configuration and Figure 12 shows the maximum bank angle excursion reached during the flight. Although 30 degrees of bank by itself is not a safety factor and bank angle alerts are announced at 45 degrees, the checklist calls for the bank angle to be limited to 15 degrees for flight control problems.

The number of excursions beyond thirty degrees is an indication of the workload and task saturation that was evident in the single pilot configuration and the reduced level of safety that resulted. The mean increased from four excursions in two-crew to 11 excursions in SPO. Since the RCO configurations are similar to the two-crew configuration after the resting pilot is back on the flight deck, there is minimal difference between two crew and RCO data. Interestingly enough, the maximum extent of the excursions is a better indicator of overall safety. There were small differences in the mean between the crew complement configurations and many of the worst excursions occurred right after the failure while the pilots were trying to get the aircraft under control. The large maximum in the single pilot configuration was somewhat due to the one pilot that departed controlled flight.

Figure 11. Number of Excursions beyond 30 Degrees

Figure 12. Maximum Bank Angle Achieved
Conclusions

Although the data analysis from this test is ongoing, the paper reflects the analysis for one non-normal scenario, and it supports the conclusion that anything less than two crew members will require significant redesign of automation and increased levels of automation support. Time to troubleshoot difficult problems not only increased but increased five-fold. Diversion decisions were negatively impacted, and safety of flight was compromised. Workload increased to unacceptable levels. These problems were particularly amplified because of the control problems induced by the rudder trim runaway.

The failure effects detailed here required significant skill, and the standard checklist only provided steps to do no harm and to get the aircraft on the ground as quickly and safely as possible. These facts showcase how failures of aircraft systems require pilot interaction for continued safety of flight.

These data also showed that the levels of experience and ingenuity displayed by the crews played an important role in the outcomes. Some interventions were straightforward (sharing time in control of the aircraft because of holding 40 lbs of rudder input) whereas others were more clever (e.g., flying with split throttles, flying with a partial sideslip configuration, jamming a shoe under the rudder pedals). The challenge will be to identify and develop technology requirements from these data for increasingly autonomous systems that might assist future operations. Autonomy – machine intelligent systems with the capability for learning – will be critical to react to these dynamic and unplanned situations. Human-autonomy teaming will also be critical since these systems must respond in intuitive and predictable manners to help and not hinder what can be hazardous conditions.

References


Acknowledgements

This work was sponsored by NASA’s Airspace Operations and Safety Program, Safe Autonomous Systems Operations project, led by Parimal Kopardekar, Project Manager.

This work was conducted as part of an Interagency Agreement, No. IA1-20800, between the National Aeronautics and Space Administration (NASA), Langley Research Center, and the U.S. Department Of Transportation, FAA. This work could not have been possible without the hard work, diligence, assistance and support of Mr. Harry Hodges (AFS-440), Mr. Doug Rodzon, and the entire AFS-440 Flight Simulation Branch who were instrumental to this research. The support of Mr. Bruce DeCleene (AFS-400) was also critical and is gratefully acknowledged.

The technical assistance of Dr. Robert “Buck” Joslin - the FAA Chief Scientist and Technical Advisor for Flight Deck Technology Integration - was critical to the simulation planning. Further, the help of Mr. Bryan Burks, Alaska Airlines, was instrumental in the development of our test scenarios.

Email Addresses

timothy.j.etherington@nasa.gov
lynda.j.kramer@nasa.gov
randall.e.bailey@nasa.gov
kellie.d.kennedy@nasa.gov

35th Digital Avionics Systems Conference
September 13-17, 2016