

Quantifying Pilot Contribution to Flight Safety during Hydraulic Systems Failure

Lynda J. Kramer¹, Timothy J. Etherington², Randall E. Bailey¹,
and Kellie D. Kennedy¹

¹ NASA Langley Research Center, Hampton, Virginia, USA

² Rockwell Collins, Cedar Rapids, Iowa, USA

Abstract. Accident statistics cite the flight crew as a causal factor in over 60% of large transport aircraft fatal accidents. 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. The latter statement, while generally accepted, cannot be verified because little or no quantitative data exists on how and how many accidents/incidents are averted by crew actions. A joint NASA/FAA high-fidelity motion-base human-in-the-loop test was conducted using a Level D certified Boeing 737-800 simulator to evaluate the pilot's contribution to safety-of-flight during routine air carrier flight operations and in response to aircraft system failures. To quantify the human's contribution, crew complement (two-crew, reduced crew, single pilot) was used as the independent variable in a between-subjects design. This paper details the crew's actions, including decision-making, and responses while dealing with a hydraulic systems leak – one of 6 total non-normal events that were simulated in this experiment.

Keywords: Flight crew error ··· Aviation Safety ··· Increasingly autonomous systems ··· Automation ··· Crew complement ··· Reduced crew operations ··· Hydraulic system

1 Introduction

Accident statistics cite the flight crew as a causal factor in over 60% of large transport aircraft fatal accidents [1]. Yet, the Air Line Pilots Association says that “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, while generally accepted, cannot be verified because little or no quantitative data exists on how and how many accidents/incidents are averted by crew actions. Anecdotal evidence suggests crews handle routine failures on a daily basis and Aviation Safety Action Program (ASAP) data [3, 4] supports this assertion but its data is not publicly releasable. Without hard data, the contribution and

methods employed by pilots to improve the safety of flight is difficult to define. Developing ways to augment and/or improve a pilot's ability to contribute to flight safety is similarly ill-defined and is hard to characterize in the absence of quantifiable data.

A joint NASA/FAA high-fidelity motion-base simulation experiment specifically addressed this void by collecting data to quantify the human (pilot) contribution to safety-of-flight and the methods used by pilots in today's National Airspace System as they 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 increasingly autonomous systems that can better support the human in the cockpit. Different crew complement configurations were tested 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 where 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 (i.e., simulating an incapacitated pilot). This paper details preliminary results and analysis of one of six non-normal events tested – a hydraulic leak in the System A reservoir.

2 Methodology

Crew complement (single pilot and crewed configurations) was experimentally manipulated during normal and increasingly challenging non-normal airline operations to quantify the pilot contribution to flight safety.

2.1 Experiment Design

The test objectives of the experiment were as follows:

- Establish “baseline” levels of performance and safety with nominal two-crew configuration as well as collect 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,
- 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 included a temporary absence, it was designated as reduced crew operations (RCO). If the condition included a permanent absence, it was designated as single pilot operations (SPO).

The independent variables were crew complement and scenario. The three crew complement configurations were: Two-Crew, RCO, and SPO. Two normal scenarios and six non-normal scenarios were flown over the two days of data collection. The non-normal scenarios were grouped into three categories (A, B, and C), with two non-normal runs in each category. Category A featured failures, initially unannounced, with autopilot available; Category B featured announced failures with autopilot available;

and, Category C featured annunciated failures with autopilot not available. Alert type and autopilot state 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). All flights were flown to landing.

Failures were triggered near top of climb or top of descent. This paper details one Category B failure - a loss of System A hydraulic system. Reference 5 provides a detailed description of the experiment design (factors, metrics, and run matrix) and details one Category C failure.

The data shown here is taken from 18 nominal Two-Crew runs, 18 nominal SPO runs, 6 nominal RCO runs (with the Captain resting), and 18 hydraulic leak non-normal runs (6 SPO, 6 RCO, 6 Two-Crew). For the RCO configuration, the non-normal started out with First Officer flying from the right seat and the Captain resting in the left seat, isolated in sight and sound from the cockpit. Two minutes after the flying pilot was alerted to the hydraulic failure, the resting pilot returned to flying duties in the cockpit. For the SPO configuration, each pilot flew from the left seat.

2.2 Participants

Thirty-six pilots (18 crews total), representing 5 airlines, participated in this experiment. Each pilot held an Airline Transport Pilot rating and was current in the 737-800 aircraft as either Captain or First Officer. All participants were male. Crews were paired by function (Captain or First Officer) and employer to minimize conflicts in training, standard operating procedures, and crew resource management techniques. Crews were instructed to bring their company's paper and/or electronic charts and 737-800 checklists with them to further reduce conflicts in training and standard operating procedures.

2.3 Simulator

The research was conducted using the B-737-800 simulator operated by the FAA AFS-440 at Oklahoma City, OK. The simulator is Level D-certified and can be used for both initial and recurrent training. The simulator, although a Level D training device, is also fitted with experimental controls, modifications, and recording capability to support AFS-440's research mission. The fidelity of the simulator and the recording capability were both critical to this research effort.

The test was set-up to replicate a normal airline operation in today's National Airspace System. An air carrier flight from Denver (KDEN) to Albuquerque (KABQ) was used. Dispatch paperwork for the flight was provided to the crews and constituted the flight release.

The simulated weather en-route contained significant areas of convective activity along the Rocky Mountain Front Range and strong Northerly winds that 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 of data collection. Weather and visibility were designed to affect any diversion decisions [5].

A live controller and pseudo-pilot(s) 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. A confederate also served as dispatcher in

the Airline Operations Center and provided communications as necessary and appropriate when radioed.

2.4 Training

No additional training was conducted for the crews as they were qualified and current B737-800 pilots and the simulator was Level D-certified.

The crews were briefed on the purpose of the experiment and received the dispatch paperwork. The crews were instructed to use their company's standard operating procedures and checklists for the entire test, including any company dispatch calls and cabin crew communications.

Prior to boarding the aircraft, the crew reviewed the paperwork and discussed the flight plan and flight conduct. Once they boarded the aircraft, the crew did a familiarization check and reviewed the simulator safety briefing. Known simulator-isms and aircraft differences were identified and discussed with the crew prior to run initiation.

The aircraft initial condition was in the hold-short of Runway 35L at KDEN with the engines running, parking brake set. The Flight Management System (FMS) was pre-loaded with the planned flight routing and the crews were asked to double check the entries. After review and confirmation of the cockpit switches/set-up 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 two-crew operations (i.e., nominal data) to which the non-normal runs flown in the RCO and SPO configurations would be compared. The nominal flight also promoted familiarity for the two-person crew interaction during the approximately 1.3 hours of flight time required for completion. This nominal flight was flown as the first run on Day 1 of data collection for each crew.

3 Results

The results shown here describe the major findings of only one of the Category B failure conditions, a System A hydraulics failure.

The 737-800 has three hydraulic systems (A, B, and Standby) that operate independently at 3000 psi. Each has a reservoir, pumps and filters. Either A or B Hydraulic System can power all flight controls with no decrease in airplane controllability. System A also provides hydraulics for landing gear, ground spoilers, alternate brakes, Engine 1 thrust reverser, Autopilot A, normal nose wheel steering, and power transfer unit. System B provides hydraulics for leading edge flaps and slats, normal brakes, Engine 2 thrust reverser, Autopilot B, alternate nose wheel steering, landing gear transfer unit, autoslats, yaw damper, and trailing edge flaps. The Standby System provides a third source for the rudder control system and a second source for thrust reversers and leading edge flaps and slats.

The hydraulic leak failure was modeled as a large leak, at a rate of 10 gallons per minute, in the System A reservoir. When the reservoir quantity dropped to less than 18.7% full, the System A hydraulics failure was annunciated to the flight crew through illumination of the: a) ENG 1 (engine-driven pump) and ELEC 2 (electric-motor-driven

pump) LOW PRESSURE lights on the forward overhead hydraulic panel; and, b) Left and Right side MASTER CAUTION lights and HYD system annunciator light on the glare shield annunciation panel. Approximately 30 seconds later, the System A Flight Controls LOW PRESSURE light illuminated on the forward overhead flight control panel and the MASTER CAUTION lights and FLT system annunciator light illuminated on the glare shield. Additionally, if the A side autopilot was engaged, it automatically disconnected and the autopilot disconnect horn would sound. The B side autopilot was still available after the failure. If not shut down in time, approximately one minute after loss of hydraulic fluid cooling, the electric hydraulic pump OVERHEAT light would illuminate. This light would remain illuminated even after the pump had been shut down until the pump cooled down. Failure Handling and Flight Path Control

For the hydraulic leak scenario, the failure occurred approximately 5 minutes prior to the top of descent during the cruise phase of flight at 36,000 ft mean sea level (MSL) while heading south toward KABQ.

Once the failure occurred, 13 out of 18 pilots/crews declared an emergency with ATC. 11 pilots/crews requested special handling (holding pattern, vectors for descent from cruise altitude to a lower altitude, and vectors for long straight-in approach) from ATC.

There were two single-pilot runs where diagnosing and attending to the System A hydraulic failure significantly affected the pilot's airplane state awareness. These two single-pilot runs appear to be the only ones in the 18 hydraulic failure runs flown where dealing with the failure directly influenced the pilot's aircraft awareness as summarized below:

- SPO-Captain configuration run: Before the hydraulic leak occurred, the pilot was cleared to descend via the SANDIA3 arrival and a preselected altitude of 9,000 feet was set to allow the aircraft to descend at the calculated top of descent. When the failure occurred, the aircraft had not yet reached top of descent and the pilot declared an emergency and requested a hold to have time to troubleshoot the problem before descending. The pilot was given vectors to a hold and engaged heading select and then flight level change, which placed the automation in autothrottle arm mode and pitch on speed for the vertical mode. The pilot had intended to select altitude hold. The pilot asked for and was given a clearance to slow down for the hold from 0.78 to 0.7 Mach and the speed reduction caused the aircraft to start a descent. After losing almost 2000 feet of altitude over 20 seconds, the pilot realized his error and requested an altitude to hold from ATC. The pilot verbalized the altitude deviation during the run, but incorrectly thought this deviation was caused by putting the autothrottles in speed mode. Post-run, the pilot recognized his altitude excursion and said "I was convinced I was in a different mode and I descended through my altitude and didn't start my hold because I was in heading select." The pilot also commented post-run "I'm a little upset with myself here for that descent. I blew an altitude and even though I'm an emergency aircraft and all that. I really like to stay out of other people's airspace. I thought that was an objectionable excursion."
- SPO-First Officer configuration run: ATC cleared the pilot to descend via the SANDIA3 arrival to 9000 feet approximately 40 seconds after the pilot was alerted to the failure of the hydraulic system and the failed side autopilot discon-

nect horn sounded. The pilot silenced the horn, acknowledged the clearance, declared an emergency, and set selected altitude to 9000 feet. Autothrottles were engaged and the flight directors were still providing lateral and vertical corrections to the FMS-programmed path as the pilot proceeded to troubleshoot the failure. While handling the checklists and troubleshooting, the pilot was not actively hand-flying the aircraft, but it was well trimmed and continued to follow the flight path. When top of descent was reached, the automation retarded the throttle and the aircraft started a descent at the trimmed airspeed. This action did not violate the clearance but was not what the pilot was expecting as he did not actively engage a descent or follow the flight director commands once the automation began the descent. After almost ten minutes and a loss of 12,000 feet of altitude, the pilot finally re-engaged the autopilot to the non-failed side once he reached the note in the Loss of System A checklist that the autopilot B was available. Post-run, the pilot said “I was single pilot, had to unstrap myself, trying to put the gear down. Had nobody to fly the airplane or monitor the systems. I have to totally rely on the autopilot while I’m messing around with the checklists, trying to program. It’s a good thing I had at least one autopilot or it would have been a different situation.”

3.1 Checklist Usage

Time-to-first correct checklist was used as a metric for quick and proper troubleshooting of equipment problems. For the System A Hydraulic failure, it was an alerted failure with annunciation on the flight deck that had a direct entry in the Quick Reference Handbook (QRH) with the Loss of System A checklist.

Crew complement was significant ($F(2,15)=8.66$, $p=0.003$) for time-to-start Loss Of System A checklist. Crews flying in the SPO or RCO configurations took approximately three times longer to start the correct checklist than those crews in the Two-Crew configuration (see Fig. 1). [The boxplots show the median ratings, with the 25th and 75th percentile spread in the data; the maximum and minimum values; and mean ratings (connected by a line).] There were no significant differences between the SPO and RCO configurations for time-to-start Loss of System A checklist. Only in one of the six RCO crews who experienced this failure did the pilot flying (PF) start the Loss of System A checklist before the resting pilot returned to the flight deck. Recall that a fixed delay of two minutes was implemented in this experiment before the resting pilot could return to the flight deck once summoned by the PF. Taking that delay into account the average time-to-start the loss of System A checklist was similar for the Two-Crew and RCO configurations. In the SPO case, getting access to the QRH and locating the correct checklist was significantly delayed compared to the Two-Crew configuration. Initially, the SPO crew had to hand-fly the aircraft because the failure disconnected the autopilot.

Time-to-complete the checklist was considered another indicator for safely handling failures. There were two checklists, the Loss of System A and Manual Gear Extension, to execute for the System A hydraulic failure. The time-to-complete metric for the Loss of System A checklist included time to execute all checklist items up to the deferred items in the Descent, Approach, Manual Gear Extension, and Before Landing checklists. The items in the descent, approach and before landing checklists varied by airline

carrier, but were similar in their content. The manual gear extension checklist was the same for all carriers; so, the time-to-complete metric included all of the items in this checklist.

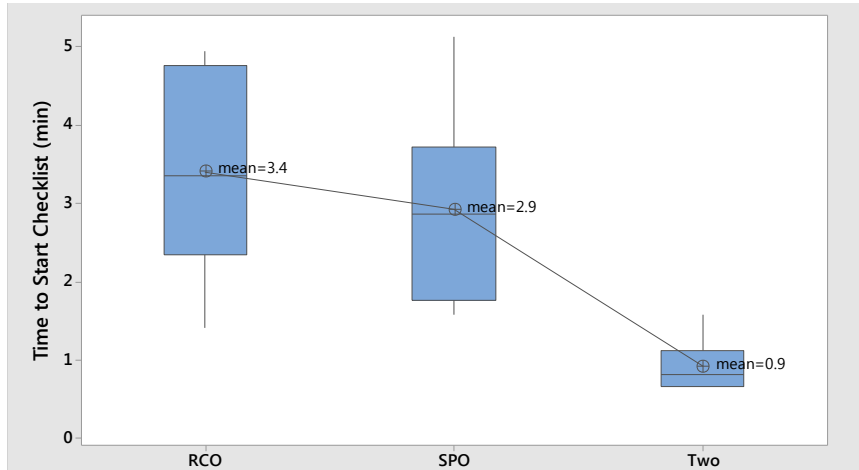


Fig. 1. Boxplot of Time to Start the Loss of System A Checklist after Hydraulic Failure

Loss of System A checklist: Crew complement was not significant ($F(2,15)=0.95$, $p=0.410$) for time-to-complete the Loss of System A checklist. The SPO configuration had the most variation for this measure as the single pilot had to simultaneously maintain aircraft control, read/execute the checklist items, communicate with ATC/dispatch, and gather weather information (See Figure 2). The overall mean time for the crews to complete the checklist was 3.3 minutes.

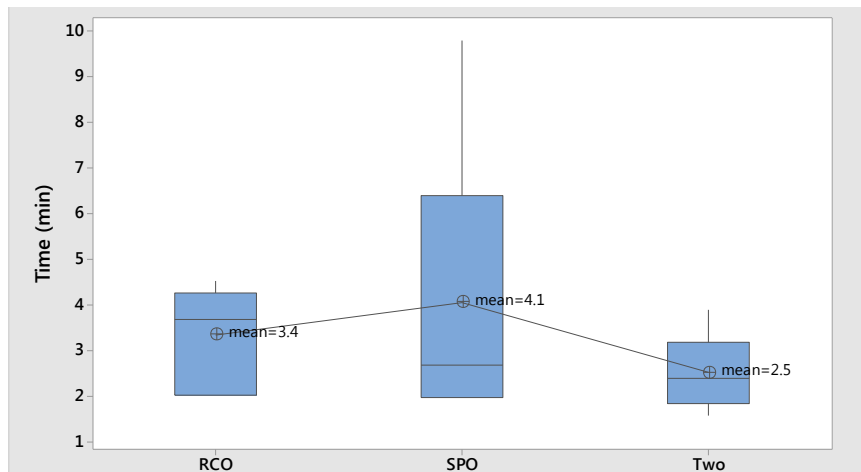


Fig. 2. Boxplot of Time Required to Complete Loss of System A Checklist

Manual Gear Extension Checklist. Crew complement was not significant ($F(2,15)=0.77$, $p=0.482$) for time-to-complete the Manual Gear Extension checklist. The SPO configuration had the most variation for this measure as the single pilot had to simultaneously maintain aircraft control, read/execute checklist items, and manually lower the landing gear (See Fig. 3). The overall mean time for the pilots to complete the checklist was 1.8 minutes.

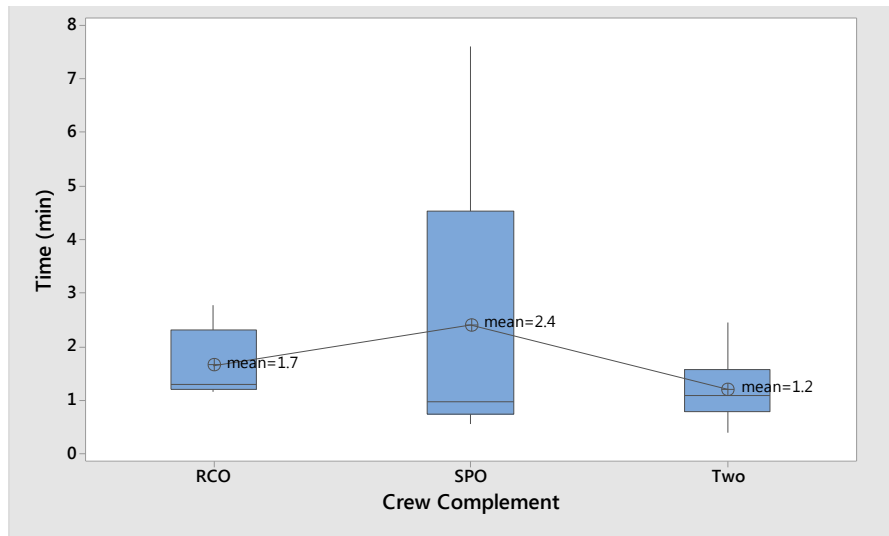


Fig. 3. Boxplot of Time Required to Complete Manual Gear Extension Checklist

The number of checklist items missed was another measure of failure handling. In 18 hydraulic failure runs, only 2 crews missed checklist items for the Loss of System A checklist. One RCO crew (with Captain resting during cruise) missed two items (turning Standby Rudder On and turning Off Hydraulic Pumps) because the First Officer did not indicate where in the checklist the resting pilot (Captain) should begin. The First Officer had the checklist out in his lap and the Captain concurred it was a Loss of System A hydraulic failure and started with the inoperative items in the checklist. The consequence of not turning the pumps off was the electric pump would continue to overheat and either burn the motor out or start a fire. Not turning standby rudder on would mean that only System B would be providing hydraulic power to the rudder. One First Officer flying in the SPO configuration missed the checklist item of turning the nose wheel steering switch to alternate. Additionally, one Captain flying in the SPO configuration missed one item, putting the landing gear lever in the down position, while completing the Manual Gear Extension checklist. This omission caused a GPWS aural warning during landing.

3.2 Diversion Decision

The test was staged to evaluate decision-making by the flight crew. A diversion decision after a failure was part of this decision-making test which tasked the pilots to consider distance to fly with the failure, the weather at each airport (KABQ and possible divert airports), and the time it took to troubleshoot the problem. Another factor that played specifically in the pilot's decision-making process for a System A hydraulic failure was that once the landing gear was manually lowered it could not be retracted which could make landing at an alternate airport impossible.

The reality is that when the hydraulic failure happened close to the top of descent, the best option was to continue a landing to the destination at KABQ. For the System A hydraulic failure, all crews, regardless of crew configuration, continued to the destination and landed safely. Santa Fe was the alternate airport for the flight but even though it was in the direct flight path, it had the same weather as Albuquerque, a shorter runway, and would require a steeper descent rate or ATC vectoring on a flight path that would be the same distance as going to the destination. Since Albuquerque, the destination airport, was only an additional 60 miles with better support facilities and the Flight Management System was already configured for flight to KABQ, Santa Fe was not considered a better alternative by any crews for the hydraulic failure runs.

3.3 Workload

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 [6]. The overall score results of this measure were examined to investigate task load variation.

Independent analyses revealed no significant ($p>0.05$) differences between the nominal runs and hydraulic failure runs for either the PF or pilot monitoring (PM) TLX ratings. For the hydraulic failure runs, pilots rated their overall workload as being low to moderate, as reflected in the PF (median rating=38) and PM (median rating=26) TLX ratings.

There were no significant ($p>0.05$) PF workload differences for crew complement. Single pilot operations were rated as having moderate workload, while crewed operations were rated as having low to moderate workload (see Fig. 4).

3.4 Safety-of-Flight

Perceived level of safety was self-assessed using a Likert type scale from 1-7, where 1 was completely acceptable and 7 was completely unacceptable.

An ANOVA revealed significant differences ($F(1,34)=7.78$, $p=0.009$) between the nominal runs and hydraulic failure runs for PF Perceived Safety of Flight ratings. Figure 5 illustrates these differences, where from an overall perceived level of safety for the PF compared to normal flight, this failure was difficult for some pilots as indicated by the large spread in data. Figure 6 shows the PF rating for each crew complement configuration. Pilots viewed the safety of this failure as unacceptable during single pilot operations (median rating=5.0) where the pilot had to simultaneously maintain

flightpath control, communicate with ATC/Dispatch, perform checklists, and eventually manually lower the landing gear. PF ratings indicated safety of flight was acceptable for this failure when there were two pilots to attend to it.

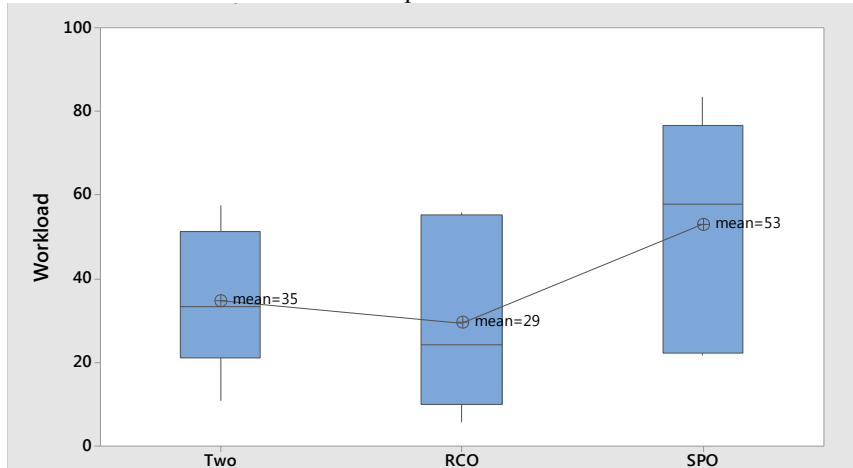


Fig. 4. Overall TLX Ratings for PF Hydraulic Failure Runs by Crew Configuration.

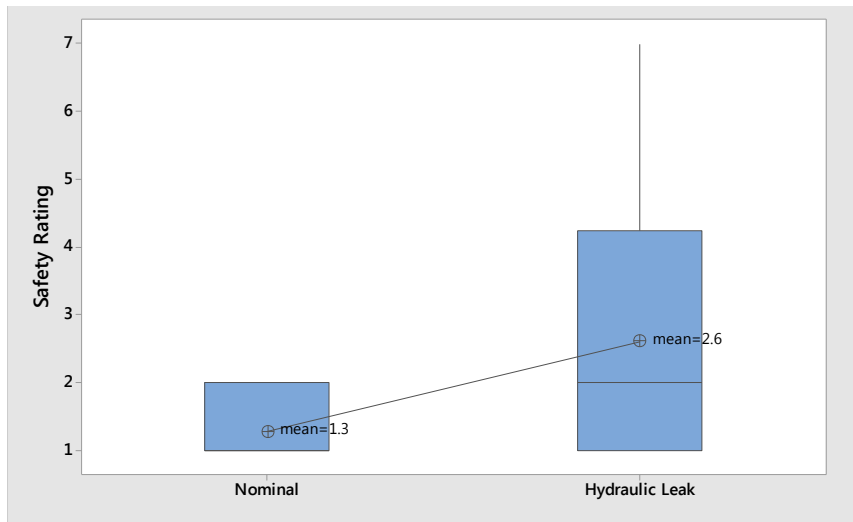


Fig. 5. Perceived Safety of Flight Ratings for Pilot Flying Nominal and Hydraulic System Failure Runs Collapsed Across Crew Configuration.

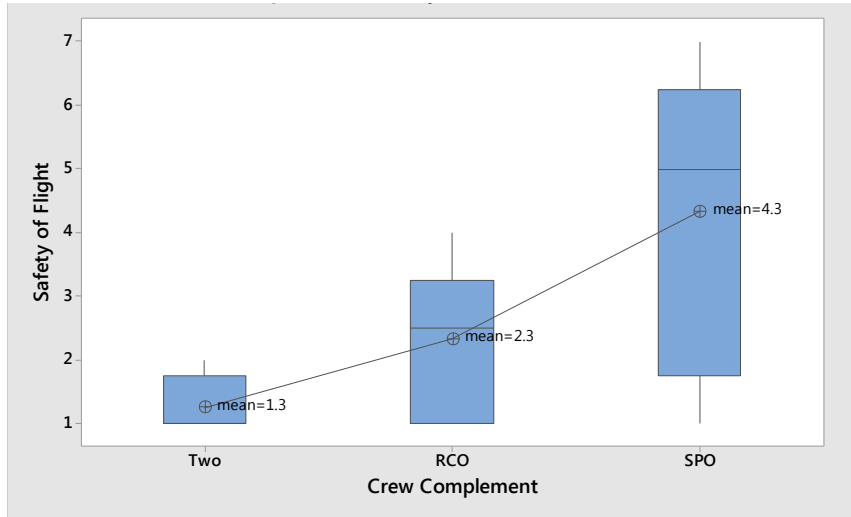


Fig. 6. Perceived Safety of Flight Ratings for Pilot-Flying Hydraulic System Failure Runs by Crew Configuration.

4 Conclusions

This paper reflects the analysis for one non-normal scenario out of six evaluated, a hydraulic system failure, 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 for a single pilot to troubleshoot and attend to failure of a primary hydraulic system increased three-fold and safety of flight was compromised. Workload increased to moderate levels during single pilot operations. The hydraulic failure occurred right before top of descent and initially caused the autopilot to disconnect. Two crews while flying single pilot lost airplane state awareness during the initial high workload phase of communicating the problem to ATC, finding the pertinent pages in the quick reference handbook, and trying to hand fly while configuring the aircraft. While not critical for this failure, loss of altitude awareness can quickly become catastrophic without the cross-check of a second pilot.

A potential latent failure was observed in the resting pilot re-engagement. The shared event of watching the failure develop, when cautions and warnings are observed, and the fact that the cautions are already reset before the resting pilot returns to the flight deck limits the shared knowledge of the resting pilot. Additionally, if the pilot flying has already opened the checklists and does not explicitly state what has already been done, the shared knowledge is further degraded. A resting Captain assumed that the First Officer had completed the checklist and concentrated on inoperative items without verifying correct aircraft configuration and missed a number of checklist items. Although mostly benign in this failure, the safety margins were decreased for the rest of the flight since some of the flight control redundancy was missing.

This failure occurred near top-of-descent; diversion to an alternate airport was not necessary. Workload and performance may have been optimistic considering the location of the failure.

Data analysis of the nominal runs and six failure runs is being used to establish quantitative baseline levels of performance and flight safety during nominal two-crew operations. This nominal data are being used to assess performance and safety decrement in reduced crew or single pilot operations using current-day flight deck design and certification. The nominal data are also being employed to identify and develop new applications and technology requirements for increasingly autonomous systems to assist pilots during dynamic and unplanned situations and perhaps future operations with two-crew, reduced crew, or possibly commercial single pilot operations.

Acknowledgments. 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.

5 References

1. Li, G., Baker, S.P., Grabowski, J.G., and Rebok, G.W., "Factors Associated with Pilot Error in Aviation Crashes," *Aviation, Space, and Environmental Medicine*, Vol. 72, No. 1, January 2001.
2. Air Line Pilots Association, "Air Line Pilots Association White Paper on Unmanned Aircraft Systems," April 2011.
3. Federal Aviation Administration Advisory Circular: Aviation Safety Action Program, AC-120-66B, Dated November, 2002.
4. Emergency and Abnormal Situations in Aviation Symposium, LOSA Data and Emergency and Abnormal Situations http://human-factors.arc.nasa.gov/flightcognition/download/EAS_Symposium_Presentations/LOSA.pdf, June 10-11, 2003.
5. Etherington, T.J., Kramer, L.J., Kennedy, K.D., Bailey, R.E., and Stephens, C.L., "Quantifying Pilot Contribution to Flight Safety for Normal and Non-Normal Airline Operations", Paper presented at AIAA Digital Avionics Systems Conference, October 2016.
6. Hart, S.G. & Staveland, L.E., "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: North-Holland, 1988, pp. 139-183.