

Quantifying Pilot Contribution to Flight Safety during Dual Generator Failure

Timothy J. Etherington
Advanced Technology Center
Rockwell Collins, Inc
Hampton, VA, USA
timothy.j.etherington@nasa.gov

Lynda J. Kramer
Crew Systems and Aviation Operations
Branch
NASA Langley Research Center
Hampton, VA, USA
lynda.j.kramer@nasa.gov

Kellie D. Kennedy
Crew Systems and Aviation Operations
Branch
NASA Langley Research Center
Hampton, VA, USA
kellie.d.kennedy@nasa.gov

Randall E. Bailey
Crew Systems and Aviation
Operations Branch
NASA Langley Research Center
Hampton, VA, USA
randall.e.bailey@nasa.gov

Mary Carolyn Last
Crew Systems and Aviation
Operations Branch
Analytical Mechanics Associates,
Inc.
Hampton, VA, USA
mary.c.last@nasa.gov

Abstract—Accident statistics cite flight crew error in over 60% of accidents involving transport category aircraft. 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 unavailable were also recorded for comparison against the nominal two-crew operations.

This paper details diversion decisions, perceived safety of flight, workload, time to complete pertinent checklists, and approach and landing results while dealing with a complete loss of electrical generators. Loss of electrical power requires pilots to complete the flight without automation support of autopilots, flight directors, or auto throttles. For reduced crew complements, the additional workload and perceived safety of flight was considered unacceptable.

Keywords— *flight crew error, aviation safety, increasingly autonomous systems, automation, automation surprise, crew complement, reduced crew operations, single pilot operations, total electrical failure, raw data approach, hand-flying skills*

I. 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 crew-members 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. This paper details preliminary results and analysis of one of six non-normal events tested – a failure of both engine driven generators.

II. 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.

A. 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 announced 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 C failure that featured a dual AC (alternating current) electrical generator failure. Also, the Auxiliary Power Unit (APU) was inoperative to increase workload for some scenarios and the APU was listed as INOP in the flight release and appropriately tagged in the flight deck. Reference [5] provides a detailed description of the experiment design (factors, metrics, and run matrix) and details one other Category C failure.

The data shown here is taken from 18 nominal Two-Crew runs and 18 dual electrical failure non-normal runs (six SPO, six RCO, six 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 signaled for the return of the resting pilot, the resting pilot returned to flying duties in the cockpit. For the SPO configuration each pilot flew from the left seat.

B. Participants

Thirty-six pilots (18 crews total), 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 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.

C. 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 designed 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 contacted.

D. 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.

E. Procedures

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 as they would on any revenue flight.

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 unique items and aircraft differences were identified and discussed with the crew prior to run initiation.

The aircraft initial condition for the nominal run was in the hold-short of Runway 35L at KDEN with the engines running, parking brake set and aircraft configured for takeoff. 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) compared to the non-normal runs flown in the RCO and SPO configurations. 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 always flown as the first run on Day 1 of data collection for each crew.

III. RESULTS

This paper details the results for one of the six failures studied, a dual electrical generator failure. Results from the other five scenarios may be found in [5, 6, 7, 8, 9]. The dual electrical failure occurred during the cruise segment of the flight approximately 15 minutes after starting the run.

This failure simulated a complete loss of AC generators coupled with the APU inoperative. Electrical failures are particularly challenging to

problem solve as they often cause multiple failures of ostensibly unrelated systems. Based on dynamic current draw and load shedding, systems become disabled seemingly at random and produce subsequent failures that are temporally disconnected from the initial failure event.

When the generators trip offline, the cockpit becomes dark and several alarms trigger before the emergency power re-enables a subset of flight instruments. The master caution panel is activated with up to eight lights since the majority of systems have redundant parts that are electrically powered. There are twelve main warning lights in total so this illustrates the difficulty in troubleshooting electrical failures. There are six additional lights that indicate loss of both electrical generators. Then there are an additional eighteen secondary lights on the overhead panel that indicate fuel and hydraulic pumps are off, window and pitot heat are off, various cooling fans and associated equipment are off, and the pressurization system automatic mode has failed. None of the secondary issues are covered in the primary checklist and require the flight crew to understand which of the eighteen lights on the overhead can be managed and which ones to ignore that are just part of an electrical failure.

In this electrical failure, the battery supplies only enough emergency AC power to operate a subset of the flight instruments for a minimum of 30 minutes. Some airlines exercise an option for a second battery that will supply emergency power for a minimum of 60 minutes.

Many systems are offline during a dual electrical failure. Traffic, terrain, weather radar, autopilot, auto throttle, flight director and right side displays do not operate. Left side primary flight display and engine displays are available but the pilot has to fly manually and operate the throttles for the remainder of the flight. One of the flight management systems continues to function. The electric fuel pumps fail during a dual AC power failure and a checklist note indicates “high altitude thrust determination and flame out may occur.” Pilots must descend below FL250 if all electric pumps are not operating.

Air-conditioning units called PACKS must remain on-line to provide pressurization but load shedding can cause high temperature problems which can cause the PACKS to randomly trip off-line. The

pressurization system AUTO FAIL light illuminates if the PACKS trip and the system cannot maintain pressurization. The outflow valve will close during this process. The only way to restore pressurization is to switch to manual mode and to bring a PACK online. The PACKS will continue to trip offline because of cooling issues and will need to be reset each time they trip.

An additional complication for crews requesting help from ATC or dispatch was the loss of one radio. The crews that did discuss the possibility of calling dispatch decided not to because they would have been unable to monitor ATC communications during that time.

The boxplots in the following figures 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).

A. Failure Handling and Flight Path Control

Although total electrical failures are not common, they do happen [10]. A dual electrical failure requires pilots to hand fly the aircraft to landing including the requirement to conduct a raw data approach in the weather conditions encountered at that airport. The flight director is also failed so the pilots did not have any command guidance. Airspeed and altitude references (bugs) are not displayed, so the expected reminders of critical values were not available. A small tick mark is displayed on the airspeed tape automatically during approach and is a reminder of 20 knots above the approach reference airspeed. If selected on the flight management system (FMS), a green REF cue is provided. All but two crews flew 20 knots above the approach reference speed all the way to touchdown.

Descending from 36,000 feet within 30 minutes requires nearly constant descent. Eighty percent of the crews diverted to Denver or Colorado Springs, which meant they flew in significant weather and turbulence the entire time. Only the pilot on the left side had flight instruments and he was required to fly the entire time to landing. Thus, the pilot flying was already fatigued by the time they had to fly the raw data ILS approach. One crew experienced an overbank during the descent portion.

All but one crew, an SPO crew, experienced a cabin depressurization and no crew effectively managed the cabin altitude during the descent

portion. No crew reset the PACKs nor did they understand that a PACK needed to be online for manual pressurization mode to work.

Flight path control during the approach was marginal for all crews. Although crew complement was not statistically significant for glideslope tracking ($F(2,17)=1.12, p=0.352$) there was operational relevance for this measure as shown in Fig. 1, where 0.33 degrees represents one dot of deviation on the glideslope scale. In a non-emergency, one dot would procedurally require a go-around below 1,000 feet. Localizer tracking, airspeed control, and managing descent rate data all showed similar trends. For SPO, success depends only on the individual pilot. For Two-Crew and RCO configurations, the availability of the pilot monitoring to coach the pilot flying during the maneuver increased success. SPO crews had the greatest variability and the glideslope mean approached one dot. By the time crews attempted the approach, Two-Crew and RCO configurations were functionally the same. Both pilots were engaged for over 20 minutes in shared operation of the aircraft and they both were involved in setting up and conducting the approach and landing. Their approach tracking performance should have been equivalent as well, yet significant variability and degraded performance was observed for RCO and the performance degradation continued all the way to landing. One SPO crew was unstable within 500 feet of the ground and should have conducted a go-around but did not. Two SPO crews did go around, one because he did not realize the flap indication was not powered until the middle of the approach and one because they were so far off in tracking that they did not have a reliable indication (full scale deflection of the instrument).

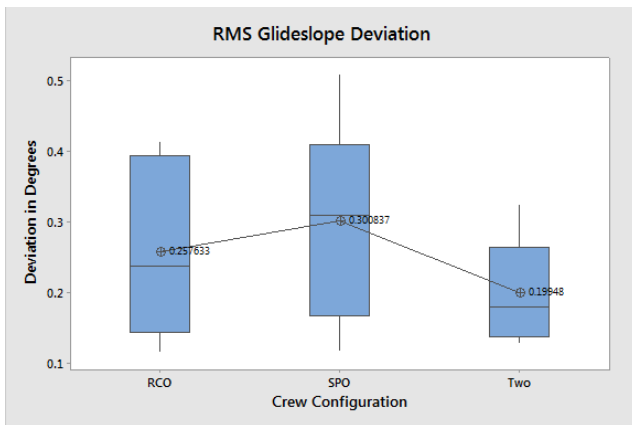


Fig. 1. RMS Glideslope Deviation by Crew Configuration

The entire flight is often judged by the landing. Table 1 shows landing performance parameters separated by crew complement. The B-737-800 has a gear distance of 16 feet, so a distance from centerline greater than 67 feet would place a tire off the edge of the runway. Landing criteria from the FAA Air Transport Pilot practical test standards is less than 3,000 feet from the end of the threshold, on centerline with airspeed no greater than 20 knots faster than the reference speed. Stabilized approach criteria are speed within five knots of the reference speed, sink rate less than 1,000 feet per minute and in a position to land without significant maneuvering. Jet aircraft are certified for landing up to 600 feet per minute without damage and sink rate at touchdown must be less than this value to ensure no damage to the aircraft. Distance from threshold for runways other than Denver is not reported due to simulator issues.

TABLE I. LANDING PERFORMANCE BY CREW CONFIGURATION

Crew Complement	Diversion Airport	Distance to Centerline	Distance from Threshold	IAS at Landing	V/S at landing	Off Runway	Loc	Gls
SPO	Denver	21	1067	218	261	Yes	0.8	1.5
SPO	Denver	13	1367	148	279		0.2	0.8
SPO	Denver	9	2934	164	277		0.3	0.4
SPO	Denver	19	897	161	343		0.5	1.1
SPO	Denver	4	1717	138	76		0.9	1.1
SPO	Denver	9	694	158	278		0.3	0.6
RCO	Denver	5	767	155	396		0.3	0.5
RCO	Denver	7	683	156	270		0.4	1.2
RCO	Santa Fe	0	*	156	300		0.1	0.9
RCO	Denver	4	1190	163	391		0.4	1.2
RCO	Grand Junction	0	*	157	313		0.4	0.5
RCO	Denver	23	4248	160	637	Yes	0.8	0.3
TWO	Denver	8	523	155	389		0.4	0.6
TWO	Denver	6	1493	164	241		0.3	0.7
TWO	Denver	2	1060	158	323		0.5	1
TWO	Colorado Springs	3	*	150	298		0.1	0.5
TWO	Grand Junction	7	*	151	211		0.3	0.4
TWO	Grand Junction	2	*	138	150		0.5	0.4

All landings were made at night and the landing lights were not operational due to the electrical system failure. All crews selected 30 flaps except one SPO crew that failed to extend flaps at all and landed at 200 knots. The reference speed for flaps 30 landing was approximately 140 knots. Crews normally select the intended landing flaps value with the flight management system and then set the airspeed bug (a magenta triangle) to the calculated reference speed plus a safety margin based on factors like wind gusts. During the electrical failure, the magenta bug was inoperative and a small white tick mark is displayed as normal at 20 knots above the approach reference speed.

Airspeed control appeared to be problematic but still acceptable, with almost all crews flying 20 knots above the approach reference speed all the way to landing. Only two crews, one SPO and one Two-Crew, achieved normal stabilized landing speed. The one pilot that landed at 200 knots was over the design

tire speed. One RCO landed beyond the design limit (greater than 600 fpm) of the aircraft. The fast SPO crew and this RCO crew also exited the lateral confines of the runway during the high speed portion of the rollout but managed to get the aircraft back on the runway prior to bringing the aircraft to a stop.

B. Checklist Usage

Time to first correct checklist is a metric for understanding and solving the problem presented. All crews recognized the dual electrical failure and many reset the generators before opening the quick reference handbook. Timely reference to the checklist was important because the checklist contained the high altitude thrust determination notice, land at the nearest suitable airport, avoid icing conditions, and 30 minute battery power limitation.

Fig. 2 shows time to start the loss of both generators checklist. Crew complement was significant ($F(2,16)=20.36, p<0.001$) for both RCO and SPO as compared to Two-Crew for this measure. The large variability in times for the SPO configuration highlight the difficult task of hand flying the aircraft while deciding where to divert, and to find, open, and read the checklist. Other data from this study for failures that left the automation available (Category A and B failures) indicated a two-minute difference between Two-Crew and RCO crew configurations [6, 7]. This Category C failure denied automation for the remainder of the flight. Here, the mean difference is almost four minutes, pointing to additional difficulties for the RCO crews.

C. Diversion Decision

Dual electrical failure required a timely divert to an alternate airport due the nature of the failure and the fact that the aircraft had limited flight time on emergency power. Weather was a divert decision factor for a number of reasons. Weather radar was inoperative so crews needed to rely on ATC to keep them safe from convective weather. They no longer had icing protection on windows or air data devices. The checklist warned them to avoid icing conditions. The lower the ceiling during the approach, the longer the crew would be required to fly on instruments and as the aircraft gets closer to the ground the instruments become much more difficult to fly and pilots tend to chase the indications. Time to divert was significant ($F(2,17)=4.23, p=0.035$) for crew

complement, but time pressure influenced the quality of the divert decision.

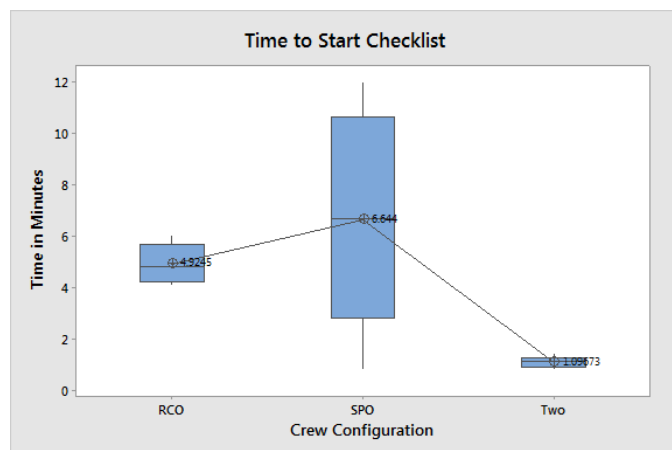


Fig. 2. Time to Start the Loss of Both Generators Checklist by Crew Configuration

Time to divert (Fig. 3) shows that pilots handling this emergency in the SPO configuration quickly made a decision to divert but all six SPO crews diverted to Denver without investigation of better weather nearby. For Two-Crew, the decision was delayed but three of six queried ATC for better weather and three of the six landed at an airport other than Denver. Note that only the Two-Crew configuration found the best diversion airport, Grand Junction, with 50% of Two-Crew making that decision. For the RCO crews, the decision on average was delayed by approximately the two minutes before the resting pilot became reengaged. Two RCO crews asked for good weather and two crews landed at an airport other than Denver, but one crew decided to fly to the filed alternate of Santa Fe, which produced a time of 48 minutes to land. This crew took the longest time to make a divert decision, 25% longer than the next longest time and three times longer than the mean time to divert.

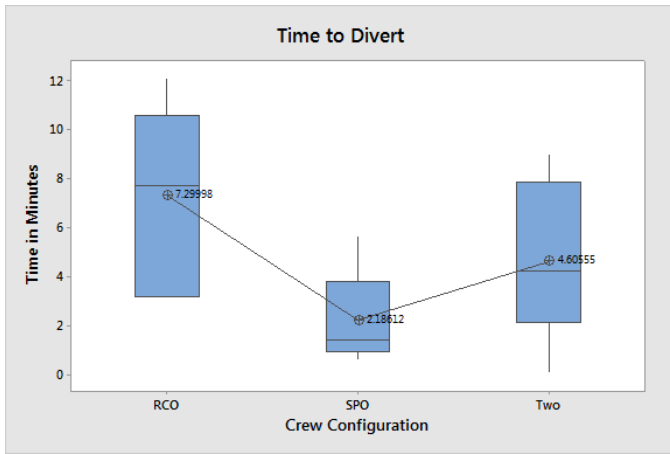


Fig. 3. Time to Communicate Diversion Decision to ATC by Crew Configuration

Fig. 4 shows the time to check weather but only some crews asked for additional weather information. The lower extent of clouds at Grand Junction was 10,000 feet, 1,000 feet for Santa Fe, and 500 feet for Denver and Colorado Springs. This meant that effectively the pilots could have been “visual” below 500 feet with plenty of time to line up on the runway. These conditions also limited the length of time in a difficult instrument segment when the instruments are very sensitive and even more difficult to fly. Two-Crew took the shortest amount of time to check weather and consistently found better weather for the diversion airport, with 50% of the crews diverting to Grand Junction.

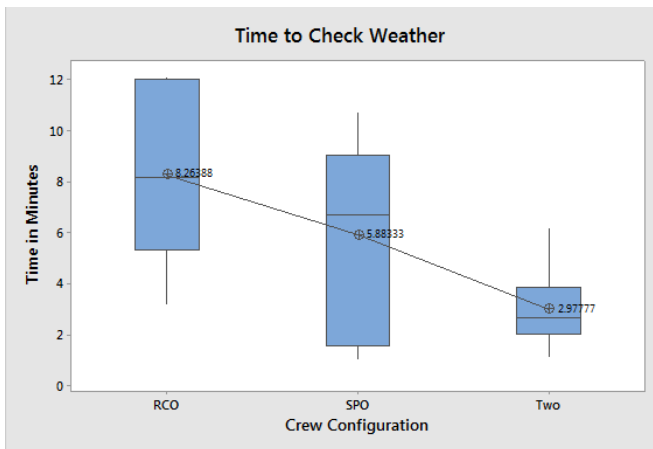


Fig. 4. Time to First Check Weather at Diversion Airport by Crew Configuration

D. Workload

The NASA Task Load Index (TLX) captured a subjective rating [11] (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. The overall score results of this measure were examined to investigate task load variation.

Not surprisingly, independent analyses revealed significant differences in TLX ratings between the nominal and dual generator failure runs for both pilot-flying (PF) ($F(1,34)=40.27, p<0.001$) and pilot-monitoring (PM) ($F(1,28)=36.78, p<0.001$) (Fig. 5 and 6). For the dual generator failure runs, pilots rated their overall workload as being moderate, as reflected in the PF (M=68) and PM (M=63) TLX ratings.

There were no significant crew complement differences for PF overall TLX ratings or PF subscale TLX ratings during a dual generator failure run. Single pilot operations were rated as having moderately high workload (M=74) while crewed operations were rated as having moderate workload (Dual [M=64], RCO [M=65]) (Fig. 7).

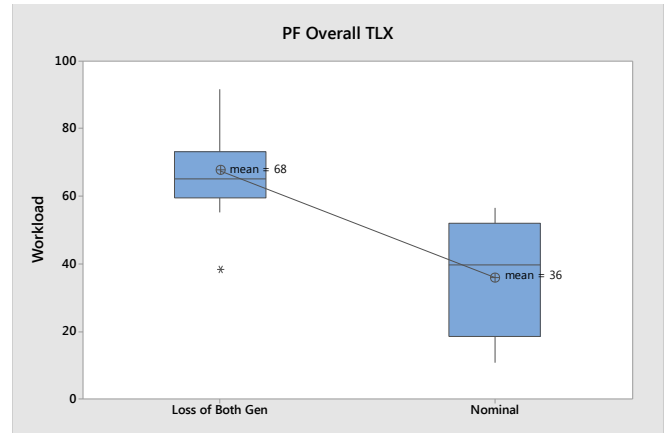


Fig. 5. Overall TLX pilot-flying ratings for dual generator failure and nominal runs

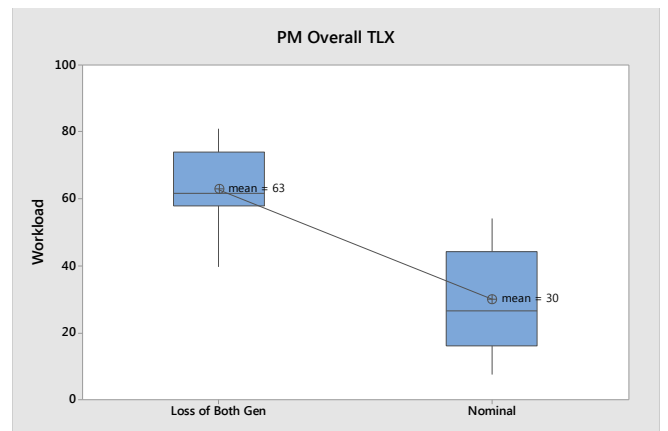


Fig. 6. Overall TLX pilot-monitoring ratings for dual generator failure and nominal runs

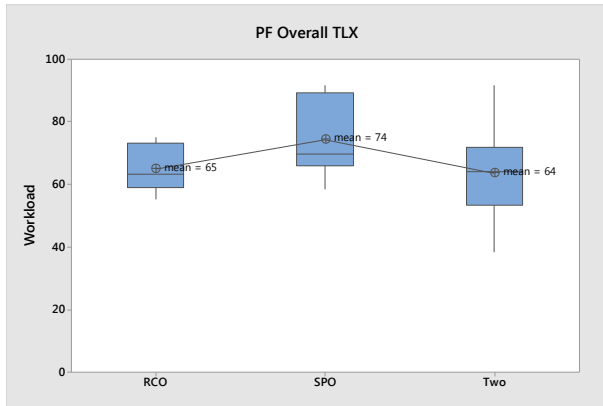


Fig. 7. Overall TLX ratings for pilot-flying dual generator failure runs by crew configuration

E. 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 for PF Perceived Safety of Flight ratings ($F(1,34)=26.57$, $p<0.001$) between the nominal runs ($M=1.3$) and the failure runs ($M=3.4$) (Fig. 8). As shown in Fig. 8, there was high variability in the overall perceived level of safety for the PF compared to normal flight. An ANOVA showed significant crew complement differences ($F(2,15)=4.20$, $p=0.036$) for PF perceived safety of flight ratings (Fig. 9). Pilots viewed the safety of this failure as unacceptable during SPO ($M=4.8$) where the pilot had to simultaneously maintain flightpath control, communicate with ATC/Dispatch, and perform checklists.

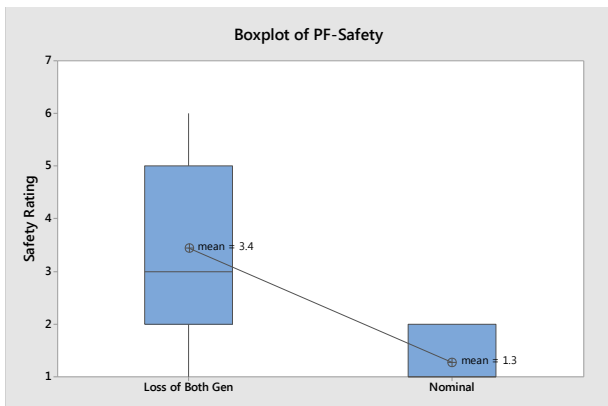


Fig. 8. Perceived safety of flight ratings for pilot-flying dual generator failure and nominal runs collapsed across crew configuration

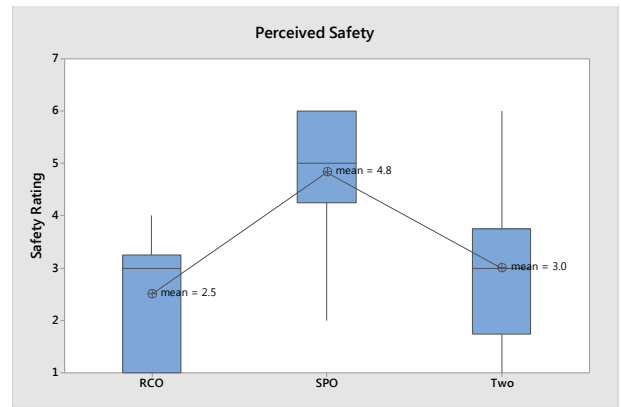


Fig. 9. Perceived safety of flight ratings for pilot-flying dual generator failure runs by crew configuration

IV. DISCUSSION

The failure was initiated at cruise altitude over the mountains and between two proximate airport landing options. Mountainous terrain limited both the number of suitable airports available and immediate descent options. If crews significantly delayed the decision to divert, Colorado Springs was further south and became an attractive option. Any crew that proceeded to the destination or the filed alternate would exceed the 30 minute battery time limit and risk significantly reduced instrumentation. Once the battery was exhausted only the integrated standby flight instrument would remain. This display provides required critical flight data but at a reduced size which increases difficulty of both normal flying and flying an already difficult ILS approach. This standby instrument runs on a separate battery with 150 minutes of life after an electrical failure. Ultimately, this failure results in a loss of all instruments and the ability to control the aircraft in instrument conditions. The majority of the crews (80%) quickly decided to divert to Denver, often before executing any checklists. SPO crews made the decision to divert almost immediately.

The weather was crafted where the most obvious diversion airport was the departing airport, Denver, but it had significant weather issues including thunderstorms, turbulence, and lower ceilings. A visual approach airport was provided in Grand Junction but the crews would have to remember the weather data from the flight release (multiple pages of data), attempt to sort through the pages in flight, ask ATC specific questions about the nearest visual

conditions, or call dispatch and ask for help. The experimental protocol was designed that if pilots asked ATC for the nearest VFR weather or airports with good weather, ATC was instructed to tell them about Grand Junction. This did not happen for two of the early runs so the diversion for Two-Crew and RCO may have been higher. For the four crews that did receive information about Grand Junction, three of the crews did divert there. By the time the fourth crew collected the necessary information, they were proximal to Colorado Springs and it became an acceptable option.

Approach tracking performance variability highlights the need for more emphasis on maintaining hand-flying ability in general [10] and flying instrument approaches with standby instrumentation [10, 12] in particular. All crews were unstable at some point in the approach below 1,000 feet above the ground. Crews are procedurally trained to go around but they have command authority above 500 feet. Conducting a landing from a marginal but salvageable approach was likely considered a better alternative than losing the electronic flight display during a go around if the battery failed. Some crews continued the approach even though the instruments were saturated and they could no longer tell if they were a little or a lot off. Crews that engaged in effective crew resource management techniques, such as the pilot monitoring providing the pilot flying with directions to fly or early indications of deviations, were much more successful than other crews or single pilots. It is not surprising that SPO had the greatest variability with the mean Glideslope RMS error at the go around limit of one dot. Only one crew was unstable below 500 feet and crews had good centerline tracking on landing. However, one RCO and one SPO crew landed near the edge of the runway on touchdown, both resulting in runway excursions. Airspeed control was affected by the airspeed tick mark at 160 knots when the approach reference speed was approximately 140 for a flaps 30 landing, which all crews except one executed.

Sink rate was excessively high. A number of things could influence the high sink rate. First, the landing lights were inoperative and all the landings occurred at night with overcast skies. Judging the flare is difficult. Second, pilot fatigue was likely as the pilots were hand-flying for almost 30 minutes before executing a raw ILS approach and landing.

Third, the highest priority during an emergency landing is to land. The data shows three pilots, two RCO and one SPO, did not flare and landed hard enough to collapse some tires on touchdown. Even in Two-Crew and SPO configurations where a third of the sink rate runs were in the desirable category, most of them approached the limit.

Time to complete the checklist was highly variable and is often driven more by ATC interaction/interruption than crew configuration. Only one SPO crew told ATC to standby when ATC interrupted the checklist flow. No crews attempted any other checklists except for emergency landing and normal descent, approach and landing checklists including pack trips or pressurization panel failures, despite these being among the eighteen lights annunciated on the overhead panel. Most crews verbalized expected pressurization issues, but when the pressurization panel was checked during the initial troubleshooting the bleed air was still providing pressurization and cabin altitude was normal. Crews commented post-run that, in the actual aircraft, their ears would have alerted them to pressurization issues before the cabin altitude horn would sound. Crews that did try to troubleshoot the pressurization problem after the horn sounded still did not have a clear understanding of the system operation. They would understand the need to select manual mode but the outflow valve was already closed meaning the reason the cabin altitude could not be maintained was not a fault in pressurization controller but that bleed was not providing positive pressure to the cabin. Earlier detection, alerted by the ears, may have improved troubleshooting or provided the time to run the checklist. Either way, support is needed for effective troubleshooting in time-critical and high workload situations.

V. CONCLUSIONS

The data suggests that even two highly qualified crew members are challenged by some failures that can occur in normal operations. This experiment reflects the worst case scenario of a critical electrical failure. The APU was inoperative, no automation support, limited diversion due to terrain, with poor nighttime weather conditions that required a raw ILS flight for an approach and landing. But these compounding issues are exactly what differentiates a failure from an accident. If even one of the conditions above had not been present it would have

significantly reduced the stress and workload of the crew during the flight. Although we did not record data on recency of hand-flying or instrument approach currency using raw data, pilot reported number of hours or time/length of employment are not indicators of success. One recent new hire first officer had some of the best performance and the worst performance was exhibited by a high time Captain. The industry should continue to stress skill maintenance to overcome degradation. One manufacturer contends that this is a basic flying skill that all pilots should already possess and it does not require special training. Although that may be true, it is a skill that requires recent practice to be effective otherwise the automation must be designed to support the pilot throughout these critical failures.

Although a dual battery option exists, many airlines do not exercise that option. When failures occur in remote areas, 30 minutes may not be enough time and certainly not enough time for a comfortable margin. Dual battery or a ram air turbine (RAT) that automatically deploys to provide emergency power would be advantageous. Although not highlighted in this experiment, even if the APU had been available, it is not recommended to attempt a start above 25,000 feet, so there will be some period of time, even with a functioning APU where the crew is working on emergency power with limited instrumentation.

Although the statistical power of the study is limited with only six RCO crews, there does appear to be an interaction with this scenario and the RCO crew complement. Time to diversion decision, time to check weather, and approach and landing performance were all degraded compared to normal Two-Crew operations. The other scenarios studied showed that even though some issues were noticed during the reengagement process, the crew interaction after that was comparable to normal two crew. There was no detriment to a resting crew member after reengagement. In this failure, performance degradation was carried all the way to landing. Further study of the data or additional studies will be required to determine the cause of this performance difference.

Automation support in determining what critical systems are affected, what things can be reconfigured, and what systems to watch as electrical loads may be balanced is lacking. Even though the 737-800 is an updated version of the classic Boeing

737, the aircraft flight deck lacks even basic engine instrument and crew alerting systems (EICAS), synoptic pages or interactive electronic checklists. These systems, prevalent on new aircraft types, would likely have significantly reduced the workload and stress on the crews in handling this emergency.

Finally, increasingly autonomous systems being researched at NASA may provide crews with a better understanding during normal and non-normal operations. Current-day automation is strong and silent yet brittle. It forces the pilot into the pilot-monitoring role for which the human is not ideally suited, especially if there is only one human on the flight deck. Autonomy, as opposed to automation, robust and chatty, with a delegation of tasks and goals from the pilot that can be easily adaptive and adaptable is the goal. Autonomy must effectively team with the human to keep them in the loop and situation aware. The communication modalities that pilots use today on the flight deck – natural language, gesture, posture, eye gaze, etc – are required by the automation if they are to enable effective communication with humans without additional workload.

ACKNOWLEDGMENT

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.

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 Air-craft 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., Bailey, R.E., Kennedy, K.D., and Stephens, C.L., Quantifying Pilot Contribution To Flight Safety For Normal And Non-Normal Airline Operations. 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, USA, 25-29 Sept 2016.
- [6] Etherington, T.J., Kramer, L.J., Bailey, R.E., and Kennedy, K.D., Quantifying Pilot Contribution To Flight Safety During An In-Flight Airspeed Failure. 19th International Symposium on Aviation Psychology, Wright State University, Dayton, Ohio, USA, 8 – 11 May 2017
- [7] Kramer, L.J., Etherington, T.J., Bailey, R.E., and Kennedy, K.D., Quantifying Pilot Contribution to Flight Safety during Hydraulic Systems Failure. 8th International Conference on Applied Human Factors and Ergonomics (AHFE 2017), Los Angeles, California, USA, 17-21 July 2017.
- [8] Kramer, L.J., Etherington, T.J., Bailey, R.E., and Kennedy, K.D., "Quantifying Pilot Contribution to Flight Safety during Drive Shaft Failure." 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), St. Petersburg, FL, USA, 17-21 Sept 2017.
- [9] Bailey, R.E., Kramer, L.J., Kennedy, K.D., Stephens, C.L. and Etherington, T.J., "An Assesment of Reduced Crew and Single Pilot Operations in Commercial Transport Aircraft Operations." 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), St. Petersburg, FL, USA, 17-21 Sept 2017.
- [10] Air Accident Investigation Board, "Accident Report No: 2/2008", (ES/C2005/10/05).
- [11] 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.
- [12] Federal Aviation Administration, "Safety Alert for Operators SAFO 13002", Flight Standard Service, Washington, DC, USA, January 2013.