A Novel Automated Electronic Checklist for Non-Normal Event Resolution Tasks

Linskens, C.E.; Reitsma, J.P.; Borst, C.; van Paassen, M.M.; Mulder, Max

DOI
10.2514/6.2021-1320

Publication date
2021

Document Version
Final published version

Published in
AIAA Scitech 2021 Forum

AIAA SciTech Forum
11–15 & 19–21 January 2021, VIRTUAL EVENT
AIAA Scitech 2021 Forum

10.2514/6.2021-1320

A Novel Automated Electronic Checklist for Non-Normal Event Resolution Tasks

C.E.Linskens, J.P.Reitsma, C.Borst, M.M. van Paassen, and M.Mulder *
Delft University of Technology, Delft, 2629HS, The Netherlands

Non-normal event resolution in-flight can be challenging on the flight crew with increased time pressure, workload, stress. Other competing tasks impose a risk on flight safety and burdens the decision-making process. Pilots rely on checklists to aid in their effort, which in its state-of-the-art form are presented on the dedicated Electronic Checklist (ECL) display for Boeing aircraft and on the Electronic Centralised Aircraft Monitor (ECAM) system for Airbus aircraft. However, human-induced errors and limitations remain prevalent. Exploring a different approach from other research efforts, this paper proposes a novel design which assumes automated checklist handling as a viable option to reduce workload durning non-normal events. In a human-in-the-loop experiment with 12 commercial pilots, the design was compared against a reproduced Boeing 787 ECL over two scenarios. A synthetic setup was used, assuming a touch-based Boeing 737-8 flight deck combined with the Boeing 787 state-of-the-art alerting systems and displays. Results indicate significant checklist completion time reductions with the proposed design of 31.3% and 42.0% for an electrical and hydraulic failure, respectively. Experienced workload and situation awareness remained unchanged, though compressed in a shorter time frame. The novel display was positively anticipated by participants but was found to lack automation feedback.

I. Introduction

A Non-Normal Checklist provides system and operational information to the pilot and step-by-step instructions to configure the flight deck in accordance with the failure, to contain, isolate deteriorating systems, to restore system functionality, and to avoid any hazardous situations. Many of the steps require the pilot to move switches and selectors, which in ECL-equipped aircraft are tied to sensors. The sensible checklist steps, or closed loop line items, allow the aircraft to check whether the step is complete [1]. Therefore, closed loop line items would make excellent automation candidates. Automating closed loop line items would alleviate the mental and physical effort required, and consequently free cognitive resources to be concentrated on resolving the non-normal event or on other urgent tasks. Additionally, no attention shifting between different panels is required, displayed checklists are shorter, and automation has the potential to complete such tasks faster than humans. Integrating automation in the process of checklist completion would allow the ECL to better attest to non-normal situation's time pressure, spikes in workload, stress, and problem-solving needs [2][3] and perhaps already set a next step towards realising SPOs.

To date, no comprehensive study was found that assumes automation as a viable approach in non-normal event resolution tasks with today's most state-of-the-art equipment. Therefore, the research question of this paper is: how does the proposed Automated ECL (AECL) design compare against the state-of-the-art ECL in terms of workload, time requirements, and situation awareness during non-normal events?

This paper proposes automation to be applied on checklist execution, an effort only considered by one other research found [4]. Comparable approaches have been explored in other research. For example, only showing the current step [5], similar to ECAM, or reducing checklist length by showing the information through synoptics [6]. However, no other studies explored the removal of checklist steps presented to the pilot, as they have now become the automation's responsibility.

The proposed design is compared against a reproduced Boeing 787 ECL in a human-in-the-loop experiment wherein 12 commercial pilots conducted two non-normal scenarios, an electrical and a hydraulic failure. The experimental setup assumed the Boeing 737-8 systems and flight deck combined with the Boeing 787 annunciation system and displays which were all operable by touch. In a between-subjects design, each participant was assigned one of the ECL displays (ECL or AECL) to conduct both scenarios and were evaluated in terms of time requirements, experienced

*Control & Simulation, Department Control & Operations, Delft University of Technology, Delft

workload, situation awareness, decision-making, managing a secondary task, and their design acceptance. Finally, potential automation drawbacks such as complacency, automation bias, and skill degradation [7] were not considered in this research.

II. The Proposed Automated ECL

Implementing an automated solution for handling checklists requires answers to the questions of what task to automate, to what extent, and when. Tasks can be categorised under four classes, (1) information acquisition, (2) information analysis, (3) decision and action selection, and (4) action implementation [8]. By using this classification we can assign a level of automation to determine the current level of automation of the ECL and determine the preferred level of automation of the proposed design. To determine if a checklist item should be automated or not, we evaluate steps based on automatability, situation awareness & authority requirements, and time requirements to determine potential benefits or concerns if the step will be automated.

A. Current level of automation of the ECL

ECLs already adopt automation within the first three classes. Acquiring information is automated, where possible, as the aircraft can detect malfunctions. Subsequently, information analysis is covered as well, since the aircraft can integrate information input into a single or multiple EICAS messages with the associated checklists displayed on the ECL. Decision and action selection provided through the predefined checklists to get the aircraft in the correct configuration and provide supplementary information as well as flight continuation advice (e.g., divert to the nearest suitable airport, avoid icing condition, or limit flight altitude to a certain height). Note that, although it is generally advised against, pilots do have authority to override checklist steps and organise the checklist order at their priority. The action implementation – the execution of the checklist steps – is however still completely manual, although it should be noted that through autosensing the aircraft supervises the pilot's actions for closed loop line items. The proposed design focuses on automating the action implementation class.

B. Selecting Automable Checklist Items

1. Automatability

Firstly, the automatability of checklist steps is assessed. Open loop line items, by its very definition, cannot be automatically sensed by the aircraft and are therefore out of consideration. They require manual completion and confirmation by the pilot and can be recognised on the ECL by the grey box in front of a step (see Figure 1). Furthermore, conditions, objectives, and operational notes do not hold a status of completion. Instead, they exclusively provide information. As such, since there is nothing to complete, there is no possibility of automation. Deferred line items refer to a Normal Checklist (NC) affected by a completed NNC. For example, a deferred line item may describe a change in the action required to complete an NC step, add or replace an individual line item, or introduce a new NC altogether. However, the deferred line items are only to be completed whenever the NC becomes relevant, for example, the Approach NC. Consequently, they are not considered for automation.

Closed loop line items are autosensed by the aircraft and, when assuming the aircraft would be capable of moving switches and selectors, have the potential to be automated. Although automatable, the different types of closed loop line items were assessed if they should be automated.

2. Situation awareness & authority requirements

The need for building situation awareness is already integrated within some of the checklist steps itself, as they can inherently differ in authority. Instructions for certain steps may indicate 'Confirm', which requires a verbal agreement of both pilots before action is taken [9]. Such steps, due to their respective impact, are classified as higher authority. They include, for example, engine thrust lever, an engine start lever, an engine, a generator drive disconnect switch. Within the class of confirmation-requiring steps, guarded switches are on the highest level of authority since a guard protects switches before they can be moved into certain positions, in addition to the required verbal pilot agreement. Such is the case for, in example, irreversible steps, which, when effectuated, are permanent and can only be reinstalled through servicing by maintenance. Consequently, any step of higher authority is excluded from automated execution.

Table 1 Overview of what is automated and integrated in the dropdown menu by NNC step type

Step Type	Automated?	Dropdown Menu?
Condition, Objective, and Operational Note	No	Yes
Closed Loop Line Item (i.e., Autosensed Action)	Yes	No
Open Loop Line Item	No	Yes
Higher Authority Items (Confirm Line Item, Guarded Switch, Irreversible Action)	No	Yes
Timer and Calculations Item	Yes	No
Closed Loop Conditional Line Item	Yes	No
Deferred Line Item	No	No

3. Time requirements

Some steps are rather time-consuming, such as calculation and timer steps. Non-normal events can cause system performance not to be up to par or the system to become inoperative, which in the case for landing-relevant systems (e.g., brakes and flaps) may increase the required landing distance. Pilots can address dedicated landing tables in the QRH to realign expectations of the aircraft landing distance accordingly. Such calculations are automatically performed with the AECL, which displays the output and output-yielding inputs (see Figure 1 for an example). Timer steps ask the pilot to wait for a certain amount of time (generally a few minutes) and are already automatically performed by the Boeing 787 ECL by displaying the time left. The AECL will also show the remaining time as well as integrating timer steps into the automation. An overview of all step types and whether they are automated is presented in Table 1.

Accredited initial automation to the AECL is low. Automation is initiated at the pilot's discretion. Beyond the initiation stage, autonomy is higher since automation will continue until finished, unless otherwise instructed by the human operator. The pilot therefore remains to have full authority over the system. The potential drawback, however, is an increased mental workload due to the added task of authorizing the automation.

C. The Dropdown Menu

Within the non-normal menu, checklists can be expanded and collapsed by pressing the right-side arrow button (see Figure 1). The expandable checklist is referred to as the dropdown menu, consisting of two main domains: the checklist content and a row from which automation and checklist progress is handled (see Figure 1). Presented content includes conditions, objectives, operational notes, open loop line items, closed loop line items of higher authority, and landing distance calculation output. Steps are included as they either present useful information or require pilot input in order to be completed. It excludes any automated steps and other non-relevant information to the pilot. The latter refers to steps overridden by conditional line items, which through an *if else* approach affects the continuation of a checklist by overriding the set of steps no longer relevant. As a result, some checklists on the ECL may appear relatively cluttered when compared against the AECL. To avoid displaying non-relevant information, the AECL dropdown menu dynamically updates when necessary after completing a conditional line item (an example is shown in Figure 2). The aim of the AECL is thus to present pertinent content only in a neat and efficient fashion. Nevertheless, the checklist as a whole is still accessible, like on the ECL, to provide flexibility since, depending on scenario circumstances and pilot knowledge and situation awareness, reviewing the grander checklist may be desired to gain further context (see Figure 1). For example, when the pilot would like to review the NNC's deferred line items, as they are not shared in the dropdown menu. An overview of what step types are included in the AECL's dropdown menu is presented in Table 1.

D. AECL Concept

Whenever a checklist contains automatable steps, the automation button can be pressed on the left to commence the automation (see Figure 2a). With automation in progress, the operator has the possibility to stop the automation. The progress of the automation is presented with an automation progress bar, which reports the fraction of the number of steps completed through automation divided by the total number of automatable steps (see Figure 2b). Once completed, the progress bar displays 'Done', as shown in Figure 2c. Additionally, from Figure 2c, it can also be observed that the checklist content was updated with two open loop line items due to a conditional line item (see Figure 2d for reference). After completing the remaining steps, the checklist displays its status of completion through the green bar stating 'Checklist Complete' (see Figure 2e). Additionally, as shown in Figure 2f, the clear button on the right appears by which

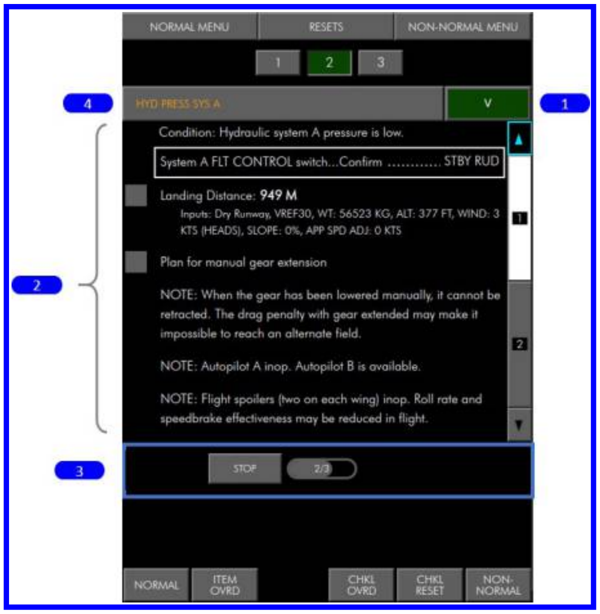


Fig. 1 By pressing the arrow on the right-hand side (1), the dropdown menu can be opened and closed. The dropdown menu itself consists of the reduced checklist content (2) and a row from which automation and checklist progress is controlled and supervised (3). When deemed necessary, the complete checklist can be accessed by pressing the checklist name (4), alike on the Boeing 787 ECL

the operator can eliminate the checklist from the non-normal menu.

III. Experiment Design

The objective of this experiment was to compare the most state-of-the-art ECL, the reproduced Boeing 787 ECL (baseline), against this research' proposed design, the AECL. In a human-in-the-loop experiment, key evaluation criteria such as experienced workload, situation awareness, and time to completion were compared in two separate scenarios on a reproduced Boeing 737 touchscreen flight deck, which adopted the Boeing 787 checklist annunciation systems.

A. Participants

In total, 14 participants volunteered to partake in the experiment. Due to the specific system knowledge requirements of the experiment, the participants are or were all professional airline pilots with a Boeing 737 type-rating. The participants were divide into two groups, the first group completed the entire experiment with the baseline ECL, and the other was presented with the AECL. The average flight hours of the participants in the baseline group was 10,150 hrs, compared to 6,800 hours in the AECL group. The average age for the baseline group was 44, and for the AECL group this was 35. The baseline group consisted out of 4 Captains and 2 First Officers were the AECL group consisted out of 2 Captains and 4 First Officers.

Due to a steep learning curve involved for learning a new display and annunciation process and to avoid scenario recognisability, the experiment used a between-subject design. Thereby, every participant was assigned a single display with which two scenarios were conducted. The order in which the two scenarios were presented was equally distributed within both groups. Furthermore, two participants were type-rated on the Boeing 787 and therefore already had experience with EICAS and the ECL. They were equally divided over the two groups as well.

B. Task and Instructions

The experimental scenario consisted of a flight departing from Almaty (UAAA), Kazakhstan to planned destination Roshchino (USTR), Russia and included one option as destination alternate, Kostanay (UAUU), Kazakhstan (see Figure 3). The locations were selected to avoid participants having previous experience with the aforementioned airports. Although GPS locations and certain airport-specific information were adopted, information about the weather,

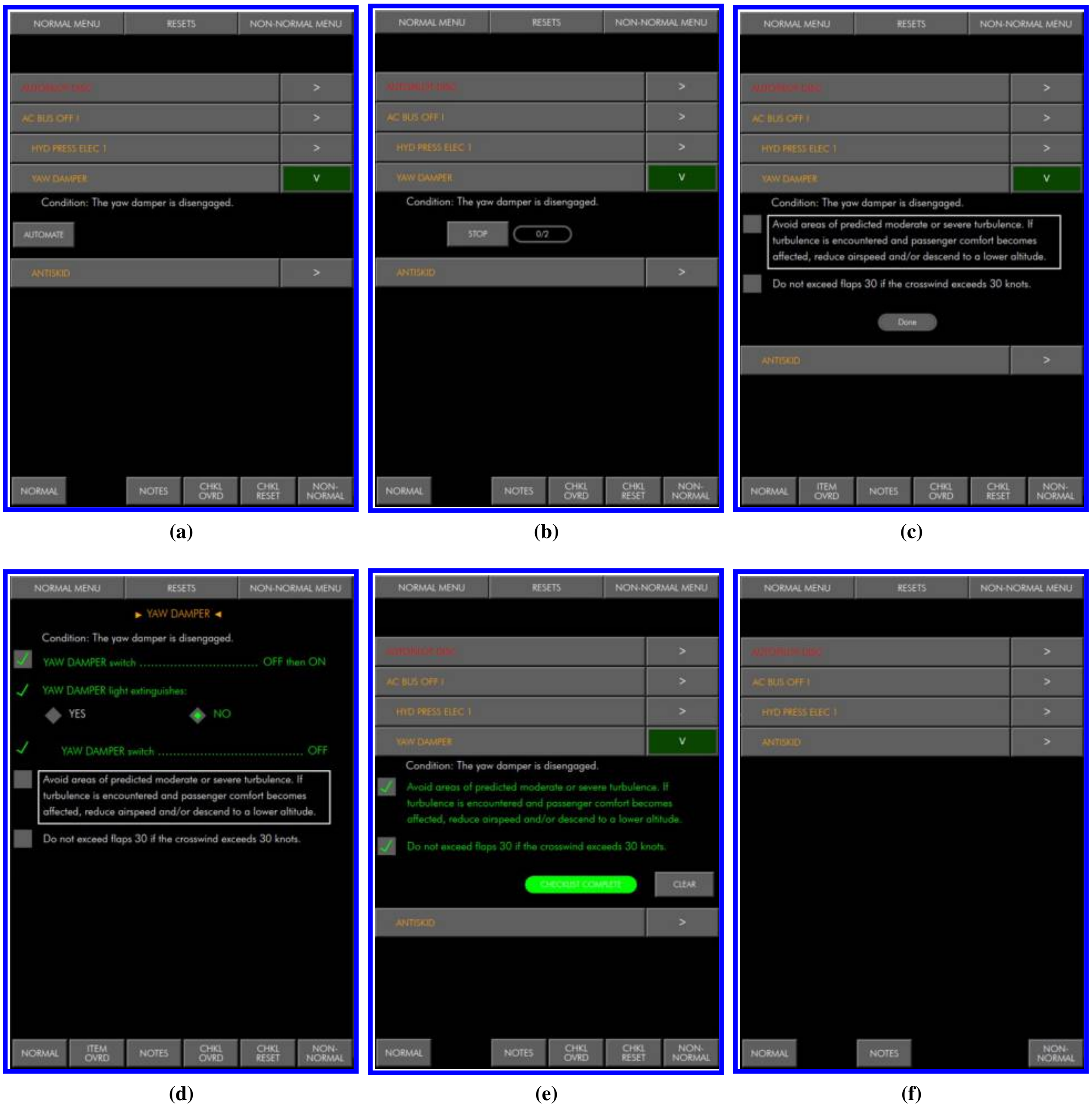


Fig. 2 Checklist automation is started in (a), in progress in (b), and finished in (c). Additionally, through conditional line items, the checklist content is dynamically updated in (c) with two more open loop line items. In (d), the AECL is compared against the reproduced Boeing 787 ECL at the current state of progress, from which it can be observed that three out of fives steps are automated and not in the dropdown menu. After completing the remaining two open loop line items in (e), the AECL displays a green bar stating 'Checklist Complete'. Finally, in (f), the checklist is now 'cleared' from the non-normal menu

runways, and approach NAVAIDs was altered to fit experiment needs and to avoid prior participant knowledge bias. Notable differences between destination and destination alternate are the distance (UAUU was closer from where the failures occurred) and the available runways and approaches. Both airports retained LOC-ILS approaches, three at USTR and one at UAUU, whereas UAUU also offered an Area Navigation (RNAV) approach. However, NOTAMs communicated that the single LOC-ILS approach runway at UAUU was inoperative at the time of flight and that no visual or circling approaches were allowed, forcing the aircraft to conduct the still available RNAV approach when diverting. To summarise, the following approaches were available per airport, at planned destination (USTR) 3 LOC-ILS and at the destination alternate only a single RNAV approach. Additionally, participants were provided with key aircraft information such as weights, a callsign, and that they were dispatched with an inoperative Auxiliary Power Unit (APU) (see Table 2 and Figure 3 for an overview of flight plan, key general aircraft, and weight information as presented to the participant).

Table 2 Information provided during experiment

Callsign	DUT 961	
Cruise altitude	FL350	
PAX	87	[-]
APU	Inoperative	
Load in compartments	1,704	[kg]
Passenger/cabin bag	9,124	[kg]
Total traffic load	10,828	[kg]
Dry operating weight	41,077	[kg]
Zero fuel weight	51,905	[kg]
Take off fuel	11,133	[kg]
Take off weight actual	63,038	[kg]
Trip fuel	7,105	[kg]
Landing weight actual	55,932	[kg]

Fig. 3 Map of the experiment flight plan

For each design, a participant completed two scenarios in which a failure occurred during the flight, an electrical and a hydraulic failure. Both scenarios assumed the same flight plan and the failure occurred approximately at the same instance, a few minutes after passing waypoint AST, from which the simulation commenced. The participants were tasked to resolve the abnormal situation when such an event would arise. Accordingly, this would require to get the aircraft in the correct configuration, and as such, all incurred checklists were to be completed. Meanwhile, the participant needed to construct a plan of approach on how to continue the flight within the context of the flight plan, wherein factors such as operational feasibility and safety were to be considered. For example, one may opt to divert to UAUU, to continue as planned to USTR, or go back to the departure airport. Together, this tests checklist handling and the decision-making process.

Also, the participants performed a secondary task. Over set intervals, five prerecorded Air Traffic Control (ATC) messages were communicated to the participant, asking to report back a particular aircraft state or element of the flight plan (e.g., flight speed, altitude, next waypoint). One of the messages corresponded to the correct callsign (DUT 961, pronounced Delta Uniform Tango Niner Six One) and the order and content of the messages were randomised to avoid learning. The secondary task increases workload demand and tests the ability to coordinate more than just one task and to what the degree a participant is tunnelled into the display.

Finally, participants assumed the Pilot Non-Flying (PNF) role and could ignore any substantial tasks generally assigned to the Pilot Flying, and were thus not concerned with flying the aircraft.

C. Independent Variables

The experiment was conducted over two dimensions of independent variables, (1) checklist support, and (2) Scenarios.

1. Checklist Support

The two ECL designs are a between-subject independent variable which compares the reproduced Boeing 787 ECL against the AECL. For both displays, participants completed two scenarios – a drive shaft failure and a hydraulic failure – under the same conditions. In contrast to the drive shaft failure, the hydraulic failure is a more commonly trained scenario for pilot training. Evaluating two of such scenarios is important since often-trained scenarios are found to be handled much better [2]. The scenarios have been verified through a level D simulator by failing the Boeing 737-8 systems as described hereinafter.

2. Scenarios

Two scenarios were used throughout the experiment namely, an electric malfunction and a hydraulic malfunction.

Drive Shaft Failure The scenario was modelled as a failing drive shaft on the left-hand side. This caused the DRIVE 1 light on the overhead panel to illuminate and EICAS to display the corresponding message. Since no APU was available to compensate, and to keep both AC transfer busses supplied, the system will close the Bus Tie Breakers (BTB)s to connect the right Integrated Drive Generators (IDG) to both transfer busses. However, the BTBs for this scenario were not functioning as expected and AC transfer bus 1 did not receive the required electrical power, causing AC transfer bus 1 to stop working. This resulted in the illumination of the SOURCE OFF and TRANSFER BUS OFF lights on the left side and two more messages on EICAS. Also, the A-side autopilot (which was engaged) was now disconnected and the autopilot disconnect horn sounded. Autopilot B, however, was still available and could be connected. By now, both the warning and caution lights illuminated on the master caution system. AC transfer bus 1 is solely responsible for powering various subsystems, which shortly failed after the loss of AC transfer bus 1. Subsequent annunciations on the overhead panel and EICAS were the YAW DAMPER, LOW PRESSURE lights for fuel pump 1 FWD, fuel pump 2 AFT, and hydraulic pump ELEC 1, TEMP PROBE, L ALPHA VANE, and L ELEV PITOT heat lights, and window OVERHEAT lights for L FWD and R SIDE. Each of which providing the associated checklist on the ECL. Additionally, ANTISKID, with an associated checklist on the ECL, is reported on EICAS as well as GPWS INOP and HIGH ALT LAND INOP which also were annunciated through the INOP light on the aft pedestal. All annunciations and checklists appeared within approximately 4 seconds after the first illumination.

Hydraulic Leak Failure The hydraulic leak in reservoir A was assumed to be relatively large, causing a loss of 10 gallons per minute . Once the reservoir quantity dropped below 18.7% of a full tank, the LOW PRESSURE lights of ENG 1 and ELEC 2 of system A on the overhead panel illuminated. After approximately 30 seconds, the system A flight controls were annunciated on the overhead, and the corresponding message was displayed on EICAS. Also, the FEEL DIFF PRESS light illuminated as a result of the hydraulic system A pressure dropping more than 25% relative to hydraulic system B. Autopilot A, the engaged autopilot, was disconnected and the horn sounded, however, Autopilot B was available and could be engaged. The master caution system illuminated both warning and caution. In case the hydraulic system was not shut down within approximately one minute, the electric hydraulic pump OVERHEAT light illuminated with the associated message shown on EICAS.

D. Control Variables

- **Concurrent task** a parallel task which required participants to respond to ATC messages throughout each scenario to increase participant workload and add realism. Also, concurrent tasks are often a constraining factor during non-normal events [2].
- **Checklists task** The checklist content was presented as per the Boeing 737-8 QRH [9].
- **Pilots** aircraft type rating, experience in flight hours, flight deck position, and current employer.
- **Flight plan** what information was presented before starting the scenarios and how this was communicated to the pilot.
- **Automation speed** the assumed time required by automation to move a switch to a certain position. For the experiment, this value was set at 0.5 seconds to guarantee a switch is in its correct position and give the flight deck ample time to recognise the new configuration before advancing.

E. Dependent Measures

The dependent measures of the experiment were as followed;

- **Experienced workload** was subjectively measured post-scenario using the Rating Scale Mental Effort (RSME) [10], a language-calibrated scale from 0 to 150 complemented with text indications to guide the participant's own ratings.
- **Situation awareness** was measured with the Situation Awareness Rating Technique (SART) [11], a post-trial subjective technique which utilises ten dimensions to obtain a single consolidated score. Both RSME and SART were only tested after each scenario to overcome unwanted intrusions and workload during the testing.
- **Performance** is appraised as a time variable, or time to completion. Two variants of completion times were assessed, the time to set the aircraft in the correct configuration following the NNCs (gross) and the time to completion when only counting time spent actually completing the checklists itself (net). Accuracy was deliberately disregarded since both ECLs only allow a checklist to acquire a status of completion when all steps are completed correctly. Furthermore, at the participant's discretion, a step may sometimes be intentionally ignored by overriding the line item, resulting in an incomparable measure.
- **Choice of airport** (destination, destination alternate, or departure) was registered, as well as the time by which such decision was made.
- **Concurrent task score** was obtained by determining the accuracy with which a participant responded to the correct callsign with the correct answer.
- **Acceptance** of both displays was assessed to identify if the design was deemed effective and suitable. Following a Crew Acceptance Rating Scale (CARS) [12] flow diagram, the participant indicated a score from 1 to 10.

F. Apparatus

The experiment was conducted in a flat panel trainer setup assuming the Boeing 737-8 cockpit and systems from the point of view of the left-positioned pilot, for this experiment the PNF. However, the 737 family does not have either the EICAS or ECL display. The two displays were taken from the Boeing 787, which is considered as state-of-the-art. This research thus adopted a synthetic flight deck, taking parts of the most advanced pieces of Boeing aircraft (EICAS, ECL, and touchscreen technology) and integrate this onto the Boeing 737-8 simulation platform.

The displays were presented on, following Figure 4, **(1)** on a 15" 4:3 XGA touchscreen, **(2)** on a 42" 9:16 UHD touchscreen, **(4)** and **(5)** on a 21.5" 16:9 FHD touchscreen, **(6)** on a 15.6" 16:9 FHD touchscreen, and **(7)** on a 19.5" 9:16 FHD touchscreen. Additionally, the overhead panel was positioned in a 25 degree inclination, similar as on the Boeing 737-8.

The location of the information was presented as follows, according to Figure 4, the ECL **(1)** was positioned between the aft pedestal **(7)** and the row of displays in front of the participant. The row contained, from left to right, the master caution **(6)**, the Primary Flight Display (PFD) **(5)**, Navigation Display (ND), and EICAS (together **(4)**). The overhead panel **(2)** was positioned above the pilot at the same inclination as in the Boeing 737-8 cockpit. Finally, the Mode Control Panel (MCP) **(3)** was placed on top of the two screens in front of the participant. All screens, and the thereon presented panels and displays, were operable by touch. The exception, however, was the MCP, which was still mechanical.

Fig. 4 Experiment apparatus

G. Training

A dedicated training scenario with made-up checklists was performed multiple times to make sure the participant was fluent in navigating the display and the touch flight deck before beginning the measurement stage. To emphasise on this, a hypothetical non-normal scenario was constructed. Herein, no logical system knowledge was required; rather, the focus was on the participant becoming affluent with any type of action required during the experiment. This would include, the various step types from Table 1, a disconnecting autopilot, and the various functions of the ECL/AECL display. During the briefing, specific instructions were communicated that, for the AECL, checklist completion was only to be performed through automation, in order to guarantee the design was used as intended. Nonetheless, it was allowed to access the checklist before and after the completion process to give the opportunity to develop context where needed.

H. Procedure

The experiment started with a technical briefing, discussing the flight deck, EICAS, the relevant ECL display (ECL or AECL), the flight plan, and the tasks at hand (primary and secondary). Subsequently, the training phase set off which was repeated until both the participant and experimenter were completely comfortable with the participant's fluency in operating the display and flight deck in order to avoid mistakes attributable to display and flight deck unfamiliarity.

After the briefing and training, two measurement scenarios were completed: the drive shaft failure and the hydraulic leak failure, each succeeded by participants indicating their experienced workload, situation awareness, and commentary on their decision rationale and thoughts on the nature of the failure. The order of the scenarios presented was equally distributed within both groups by following a Latin Square design. A short debrief was administered when both scenarios were completed, which asked participants to indicate an acceptance score and to provide feedback on the design and touch flight deck. In total, the experiment duration averaged around 3-3.5 hours per participant.

I. Hypotheses

It was hypothesised that for the AECL, when compared against the baseline ECL:

- Experienced workload decreases as a result of automation.
- Time to completion decreases. With automation, less time is required to get the aircraft in the correct configuration. Moreover, since less attention shifting is required when omitting the manual work, participants can better focus on solving the non-normal event.
- Situation awareness is expected to remain unchanged. The automated design might suffer from out-of-the-loop complications [13] in terms of perception, since part of the aircraft's non-normal configuration is no longer done manually. However, such effects are expected to be minimal and not influence results. On the other hand, situation awareness may increase because of the freed cognitive resources due to the automation, which would allow for better comprehension and projection of future status. Again such effects are expected marginal since every participant during the experiment is not constrained in time.
- The concurrent task score is expected higher due to a product of lower expected experienced workload (and thus an enhanced capability to manage other tasks) and less attentional tunnelling when not manually completing steps.

IV. Results

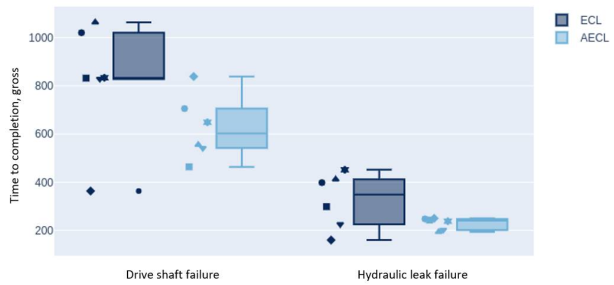
Seven participants per display completed the experiment, but data from two participants (one from each group) were removed as one did not complete the scenarios as instructed and the other because of incomplete data. Furthermore, parametric assumptions were violated for the between-subjects setup, and due to the small sample size, six for both displays, the Mann-Whitney U test is utilised where applicable.

A. Time to Completion

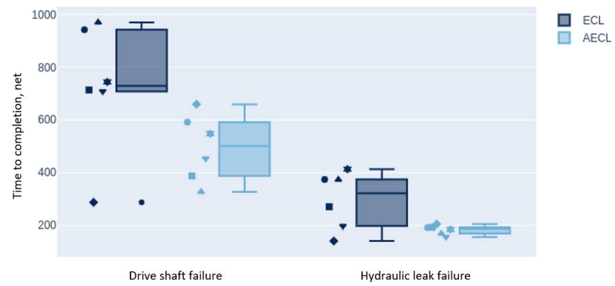
Time to completion is assessed in two ways. First, the gross value is analysed, which describes the time required to finish all checklists, and secondly, the net time to completion is considered, which considers the time actually spent with the checklists.

1. Gross time to Completion

The gross time to completion results are visualised in Figure 5a. With the AECL, the median is considerably lower for the drive shaft failure (ECL: Mdn = 832.8, AECL: Mdn = 602.2) and the hydraulic leak failure (ECL: Mdn = 348.5, AECL: Mdn = 240.4), with a drop of 27.7% and 31.0%, respectively. Statistical results, however, do not report significance for the drive shaft failure scenario ($U = 9.0$, $p = 0.087$) and the hydraulic leak failure scenario ($U = 10.0$, $p = 0.115$). This may be partially affected by one participant achieving extreme scores with the ECL display. Apart from this participant, all of the participants using the AECL achieved lower times to completion for the drive shaft failure, except the AECL diamond, which time to completion equalled the lower end of the ECL display. Similar trends can be observed for the hydraulic leak failure, where the AECL's time to completions are in the low range of the ECL



(a) Gross time to completion [seconds]



(b) Net time to completion [seconds]

Fig. 5 Time to completion results

display. Additionally, time to completion across all participants is very consistent for the AECL display in the hydraulic leak failure scenario. This is likely the consequence of the hydraulic leak failure scenario only having one substantial checklist (loss of system A), wherewith the drive shaft failure scenario 12 checklists appear.

2. Net time to Completion

The net time of completion, on the other hand, does show a significance for both the drive shaft failure (U = 6.0, p = 0.033) and the hydraulic leak failure (U = 7.0, p = 0.046). Most likely, this measure better articulates the increased time efficiency of the AECL since the datapoints of the AECL display decrease more relative to the ECL display, as can be observed when comparing Figure 5b with Figure 5a. The net time to completion medians are reduced by 31.3% for the drive shaft failure scenario (ECL: Mdn = 728.8, AECL: Mdn = 500.7) and 42.0% for the hydraulic leak scenario (ECL: Mdn = 322.1, AECL: Mdn = 187.0).

B. Experienced Workload

The subjectively indicated RSME workload per design for both scenarios is shown in Figure 6, with a higher median for the AECL display in the drive shaft failure scenario (ECL: Mdn = 37.5, AECL: Mdn = 50.0) and a slightly lower median in the hydraulic leak failure scenario (ECL: Mdn = 32.5, AECL: Mdn = 32.0). The Mann-Whitney U test reveals no significant effect in the drive shaft failure scenario (U = 15.0, p = 0.343) and the hydraulic leak failure (U = 16.0, p = 0.404). Examining Figure 6 more closely reveals that for the AECL, especially one participant indicated higher experienced workloads. Also, the experienced workload scores are relatively widely spread for all experiment conditions, perhaps with the exception for the AECL in the hydraulic leak failure, which is more condensed apart from one outlier. Nonetheless, this would indicate a large variation on an individual basis and a recurrence of participants near the extremes. The results indicate no difference between both designs in terms of experienced workload, which does not support the hypothesis.

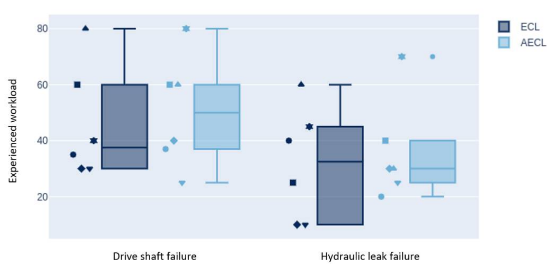


Fig. 6 Experienced workload ratings

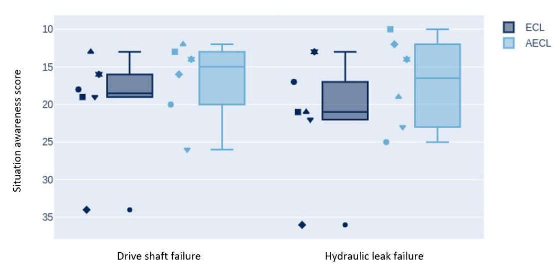


Fig. 7 Situation awareness scores

C. Situation Awareness

The situation awareness SART measurements are displayed in Figure 7 (note that the scores are from high to low). When consulting the Mann-Whitney U test, results are insignificant for the drive shaft failure ($U = 13.0$, $p = 0.234$) and the hydraulic leak failure ($U = 13.0$, $p = 0.235$). Median values are however higher for the ECL display across both the drive shaft failure (ECL: Mdn = 18.5, AECL: Mdn = 15.0) and the hydraulic leak failure (ECL: Mdn = 21.0, AECL: Mdn = 16.5) scenarios, partially driven by the outlier of the ECL display. Another interesting observation is the more defined range of situation awareness for the ECL display, as the AECL values are more diffused in both scenarios.

The statistical insignificance and data observations indicate no difference in situation awareness for both displays and would support the hypothesis that despite the introduction of automation, no significant impairment of situation awareness occurs. This includes automation side effects such as becoming out-of-the-loop. Participant comments, on the other hand, proved otherwise. It was a deliberate design choice to minimise any information shown and that participants would have context enough to understand what switches and selectors were operated by the automation. Although most participants were reasonably confident, the desire for more feedback about the ultimate automation results was unanimous. Such commentary was mentioned in the post-experiment questions, either in asking about participant trust in automation or in what could be improved on the current design.

Finally, in context of the hypothesis, it cannot be concluded to what extent automation negatively contributed due to being out-of-the-loop and how much it positively influenced participants through freed cognitive resources to be allocated to develop situation awareness.

D. Choice of Airport

Upon experiencing a failure, participants had the choice to either continue as planned, or choose to divert to destination alternate. For the drive shaft scenario, the checklists communicated to the pilot to land at the nearest suitable airport. Destination alternate is the nearest airport, and hence, the task is to determine if it is suitable. As the scenario is set up, the question is whether it is authorised to conduct an RNAV approach with the incurred failures and land safely on the runway. Other literature indicates that with the same failure, RNAV approaches are no longer approved when AC transfer bus 1 and the APU are inoperative [14]. However, as per the operating manuals for some of the airline companies of which participants took part in the experiment, the minimum RNAV requirements were not violated. Likely this is the result of the ongoing development of airline companies' risk position towards RNAV approaches. As such, both options are possible. The added distance of the planned destination is not substantially greater and would be commercially and operationally more attractive. On the other hand, in some cases, the specific checklist instruction of landing at the nearest airport would be ignored when not diverting to destination alternate.

For the hydraulic failure, the best option is to continue as planned, since after losing hydraulic system A, there is still hydraulic redundancy left with hydraulic system B and the standby hydraulic system. Moreover, both hydraulic system A and B are capable of single-handedly powering flight controls without losing controllability [15]. However, the choice of airport is one of full commitment since the landing gear has to be lowered manually, increasing deployment time, and it can, once extended, no longer be retracted. As a result, the aircraft suffers from a permanent drag penalty, making other airports unreachable after a landing attempt. Therefore, when selecting destination alternate, one becomes fully reliable on one runway and takes a more risky option.

For the drive shaft failure, four out of six participants diverted with the AECL display, whereas for the ECL five out of six chose to divert, as can be seen in Figure 8. The time required to form a decision is shown in Figure 10, where the AECL achieved lower median values compared to the ECL display (ECL: Mdn = 474.7, AECL: Mdn = 233.2). The Mann-Whitney U test indeed reveals a significantly lower decision time for the AECL ($U = 7.0$, $p = 0.046$).

Interestingly, after investigation of experiment video recordings and post-scenario commentary, none of the participants considered whether an RNAV approach was still authorised, including participants for which their current airline does not approve RNAV approaches under this failure. Post experiment, participants were asked to describe to the best of their knowledge, the nature of the failure and what drove their decision for the selected airport.

With the hydraulic leak failure, for both designs, participants diverted two out of six times, as shown in Figure 9. Again, Figure 10 reveals a lower median decision time for the AECL display (ECL: Mdn = 328.6, AECL: Mdn = 157.7), and when consulting the Mann-Whitney U test, the lower decision time is found to be significant ($U = 7.0$, $p = 0.046$). Therefore, the AECL display shows significant time reductions for both scenarios in formulating a decision when compared against the ECL display.

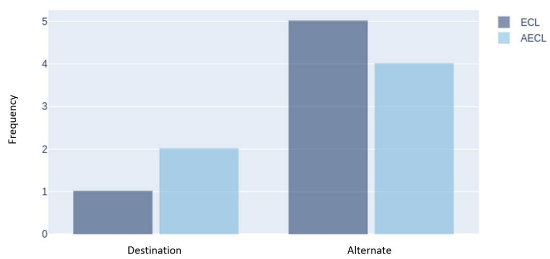


Fig. 8 Chosen airport during the drive shaft failure scenario.

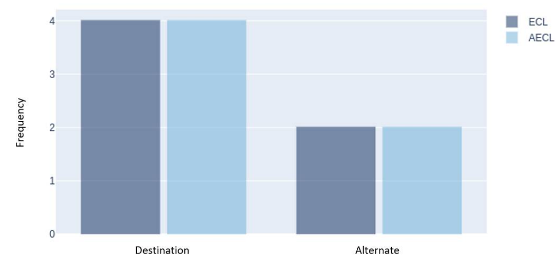


Fig. 9 Chosen airport during the hydraulic leak failure scenario.

E. Acceptance

The acceptance scores obtained through the CARS measurement are summarised in Figure 11 for the ECL and AECL display. When observing the figure, the AECL has a more apparent consensus, with five out of six results equalling 8 out of 10. The ECL shows more variability with scores mainly ranging between 7 and 9 and an outlier of 1. The outlier participant commented that the ECL negatively affected situation awareness and lacked overview. Furthermore, the AECL has a slightly higher median value (ECL: Mdn = 7.5, AECL: Mdn = 8.0). This difference, however, is found not to be significant according to the Mann-Whitney U test (U = 16.5, p = 0.431). As already mentioned, participants reported a lack of automation feedback, possibly capping the acceptance score at 8 for the AECL display.

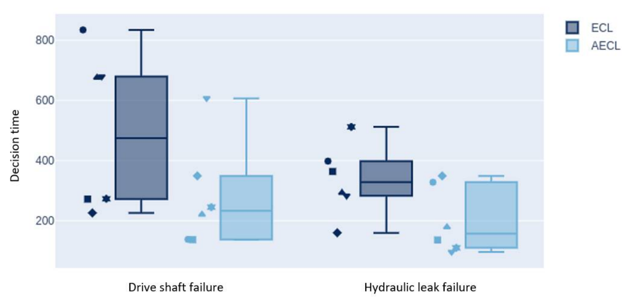


Fig. 10 Decision time.

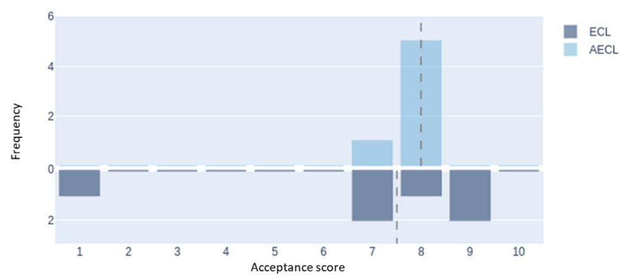


Fig. 11 Acceptance scores.

F. Concurrent Task

The concurrent task score indicates the accuracy by which a participant completed the challenge-response task. Per scenario, Table 3 shows whether the participant successfully completed each concurrent task, in which only data points after introduction of the failure were considered. Since participants had varying times of completion and the concurrent tasks were initiated at set time intervals, some participants completed more tasks than others. To negate this effect for the statistical analysis, only the concurrent tasks performed by every participant per scenario after introduction of the failure (three for the drive shaft failure and one for the hydraulic leak failure) are included. Table 3 shows little difference between both designs for the concurrent task accuracy. For the hydraulic leak failure, it can be observed that one more participant achieved a perfect score with the ECL display, which has no effect on the median values (ECL: Mdn = 100.0%, AECL: Mdn = 100.0%). Unsurprisingly, the Mann-Whitney U test reveals no significance (U = 15.0, p = 0.297). For the drive shaft failure scenario, the AECL slightly outperforms (ECL: Mdn = 83.3%, AECL: Mdn = 100.0%), but no significance was found (U = 14.0, p = 0.261).

V. Discussion

The goal of this research was to investigate the effects of adopting an automation effort to improve the ECL for non-normal situations, an attempt to achieve lower workload and time requirements, while maintaining situation awareness. Results revealed that particularly completing time of a checklists and time to reach a final decision-making were significantly reduced with the AECL.

Table 3 Correct or incorrect completion of the concurrent tasks per participant after introduction of a failure

		Drive shaft failure					Hydraulic leak failure		
	Participant	1	2	3	4	Accuracy*	1	2	Accuracy†
ECL	1	×	✓	✓	✓	67%	✓	✓	100%
	2	✓	✓	✓	×	100%	✓		100%
	3	✓	✓	✓		100%	✓	✓	100%
	4	×	✓	✓	✓	67%	✓	✓	100%
	5	✓	✓	✓	✓	100%	✓		100%
	6	×	✓	×	×	33%	×	✓	0%
AECL	7	×	✓	✓		67%	×		0%
	8	✓	✓	✓		100%	×		0%
	9	✓	✓	✓	✓	100%	✓		100%
	10	×	✓	✓		67%	✓		100%
	11	✓	✓	✓		100%	✓	✓	100%
	12	✓	✓	✓		100%	✓		100%
Accuracy, ECL		50%	100%	83%	60%		83%	100%	
Accuracy, AECL		67%	100%	100%	100%		67%	100%	

The time required for a participant to get the flight deck in the correct configuration reduced by 27.7% for the drive shaft failure and with 31.0% for the hydraulic leak failure, which converts to a time saving of 3 minutes and 51 seconds and 1 minute and 48 seconds, respectively. Counting only the time spent inside checklists further amplifies the percentage difference to 31.1% for the drive shaft failure and 42.0% for the hydraulic leak failure. The gross time to completion did not differ significantly. However, the net time to completion did, which is arguably the better measure since it directly compares the time allocated to checklists. On the other hand, the gross value does include factors such as the participant rationalising the failure and deciding upon next steps. Nonetheless, time reductions for the AECL are substantial and indicate a more adequate approach to address a non-normal event's sometimes stringent time requirements [2].

Etherington et al. [6], with the synoptics and shortened ECL approach, found comparable time reductions of 25% for a blocked pitot-static system scenario and 30% in a left hydraulic system failure scenario. Especially the latter would directly compare against the hydraulic leak failure. Interestingly, both proposed designs realised approximately 30% time reductions, despite the distinctively different approach taken. However, Etherington et al. indicated to have achieved a large variability in time reductions, whereas for this research, the time reductions appear to be relatively consistent. Additionally, when considering the net time to completion, the AECL slightly outperforms with 42% when compared against the ECL.

The percentage difference between both the gross and net values can be explained by, firstly, an arithmetic cause since time differences are now compared against a smaller absolute value. Secondly, AECL-using participants were more likely to perform concurrent processing of checklists. For example, a checklist may ask participants to wait for two minutes. In such a case, most participants chose to continue with another checklist, since automation would take care of the remaining to be automated steps. Unlike for the baseline, where the checklists were completed in isolation (not inside the menu), from which it was perhaps less stimulating to continue with another checklist. Other factors causing discrepancies could be, for example, the interrupting concurrent task.

Experienced workload was not significantly reduced as per the RSME measure. Also, the concurrent task scores do not indicate significant differences in dealing with competing tasks. However, following the discussion on the reduced AECL time requirements, it can be argued that a comparatively equal experienced workload is achieved for the AECL, but over a shorter time frame. Although the RSME scores do not support the hypothesis of a decreased experienced workload for the AECL, above explanation might hint towards an overall experienced workload reduction. This would require further experimentation and could be enforced by putting a higher time pressure on the participants. Thomas [4] did show to achieve a lower experienced workload score for higher levels of automation. However, due to the lack of challenge in the scenario itself, the practical differences were considered minimal.

Little difference between the displays was observed in the decision-making outcome, hence, the choice of selecting the planned destination or the destination alternate. In the hydraulic leak failure scenario, participants for both displays took the most appropriate decision 67% of the time, since the planned destination is commercially most attractive and ample redundancy is in place. Furthermore, for both displays 33% of participants selected a far more risky approach by committing to destination alternate, where only one runway is available and other airports are unreachable with the now non-retractable landing gear.

For the drive shaft failure scenario, 83% for the ECL and 67% for the AECL followed checklist instructions rigorously on the basis of landing at the nearest airport, which is destination alternate. Nevertheless, the exact checklist instruction was to land at the nearest suitable airport, which in context of the flight plan is encapsulated by answering the question whether the aircraft is still authorised to fly RNAV approaches with an inoperative APU and under a drive shaft failure wherein due to an additional BTB switch malfunction, AC transfer bus 1 is lost. Despite that other research indicates a loss of AC transfer bus 1 and the APU inoperative no longer approves flying RNAV approaches [16], after careful joint review by a number of participants, not for all airline companies of which participant's volunteered the RNAV approach became unauthorised. For the other airline companies, RNAV approaches were not allowed. Notwithstanding, whether correctly diverting or not, in all cases the operational consideration whether RNAV approaches were still approved was none-existent to severely limited. Plausibly, this is due to the fact that checklists do not outline RNAV approach capabilities, rather, pilots are required to ascertain in a proactive fashion whether such approaches are still approved. Similar findings of pilot unawareness of the RNAV approach with the same scenario were found in a study by Kramer et al. [16]. This calls into question the current NNC content which could benefit from better guidance on the technical feasibility of major flight components, such as approaches. To address this issue, an avenue worth exploring are ecological interface designs. This concern is however outside the scope of this research where checklist content is defined as a control variable.

Apart from decision-making outcomes, the time by which a decision was formed was significantly reduced with the AECL design, with 50.9% and 52.0% for median values for the drive shaft failure and hydraulic leak failure, respectively. First of all, such outcomes likely correlate to decreased time to completions for the ECL. Nonetheless, the much greater percentage time gain is surprising, since for example in the drive shaft failure, every participant started with a checklist that communicated to land at the nearest suitable airport. Therefore, all participants early on received this instruction, making the large difference in decision times surprising. Closer examination of video materials reveals that participants typically first consult more checklists before making a final decision, after inspection of critical elements such as fuel availability. An explanation could be that through automation participants have information available faster (lower times to completion) and only have to focus on directly relevant information, can thereby faster comprehend the situation, and hence subsequently form a decision. However, no significant evidence was found that supports or counters the second part of this interpretation.

Following the above discussion, the proposed AECL design delivers promising results. First, significant time reductions in checklist completion times (net time to completion) and decision-making were found, which would allow the pilot to better meet non-normal event time pressure [2] and deal with higher troubleshooting times found for SPO conditions [14]. Moreover, comparable experienced workload and situation awareness was observed, but the measurements were realised within a shorter time window. This could indicate a lower overall workload and faster development of situation awareness. Therefore, for future testing, it is proposed to introduce a failure for which participants experience a higher time pressure when solving a scenario. Another possibility would be to measure throughout each scenario run, but such measurements techniques can be intrusive and for that were avoided for this experiment.

Additionally, continued design iterations regarding the communication of automation outcomes are proposed. In doing so, it is expected that higher situation awareness scores and acceptance ratings can be achieved. It is, however, a trade-off in terms of time requirements as more information may slow down the operator. On the other hand, time may be gained, since, as observed for some participants during the experiment, less time is involved in verifying the flight deck on the results of automation.

VI. Conclusions

To better support pilots during non-normal event resolution tasks, this research proposed an automated Electronic Checklist (ECL) which was tested through a human-in-the-loop experiment against a reproduced Boeing 787 ECL with 12 commercial pilots when assuming the Boeing 737 systems and flight deck. Significant reductions in time requirements were found for both the tested scenarios, with 31.3% and 42.0% lower median checklist completion times and 50.9% and 52.0% lower median decision times for the drive shaft failure and hydraulic leak failure, respectively. Following this result, pilots would be better positioned an with automated ECL during emergencies wherein speed and accuracy is of essence.

Experienced workload did not significantly differ but was for the new design compressed in a shorter time frame, indicating a potential to better match the many competing tasks onboard aircraft during non-normal situations. Despite adopting automation, no significant indications of adverse effects on situation awareness were found, supported by comparable decision-making outcomes between both designs. Although initial results are promising, participants unanimously indicated a need for more automation feedback. It is proposed for next design iterations to better communicate automation outcomes.

References

[1] *787-8 Flight Crew Operations Manual*, The Boeing Company, Seattle, Washington, USA, 2009.

[2] Burian, B., Barshi, I., and Dismukes, K., "The Challenge of Aviation Emergency and Abnormal Situations," *NASA Technical Memorandum 2005–213462*, Moffett Field, California, USA: NASA Ames Research Center, 2005.

[3] Burian, B., and Barshi, I., "Emergency and Abnormal Situations: A Review of ASRS Reports," *Proceedings of the 12th International Symposium on Aviation Psychology*, Wright State University Press, Dayton, Ohio, USA, 2003.

[4] Thomas, L.C., "Pilot workload under non-normal event resolution: Assessment of levels of automation and a voice interface," *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting*, Las Vegas, Nevada, USA, 2011, pp. 11–15.

[5] Li, W.C., Cao, J., Lin, J.H., Braithwaite, G., and Greaves, M., "The Evaluation of Pilot's First Fixation and Response Time to Different Design of Alerting Messages," *International Conference on Engineering Psychology and Cognitive Ergonomics*, Springer, Cham, Switzerland, 2017, pp. 21–31.

[6] Etherington, T.J., Kramer, L.J., Young, S.D., and Daniels, T.S., "Information Management to Mitigate Loss of Control Inflight Airline Accidents," *AIAA Scitech 2020 Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, USA, 2020.

[7] Bainbridge, L., "Ironies of automation," *Automatica*, Vol. 19, No. 6, 1983, pp. 775–779.

[8] Parasuraman, R., Sheridan, T.B., and Wickens, C.D., "A model for types and levels of human interaction with automation," *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, Vol. 30, No. 3, 2000, pp. 286–297.

[9] *737 Quick Reference Handbook*, The Boeing Company, Seattle, Washington, USA, 2013.

[10] Zijlstra, F.R.H., "Efficiency in Work Behaviour – a design approach for modern tools," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 1993.

[11] Taylor, R.M., "Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design," *Situational Awareness in Aerospace Operations (AGARD-CP-478)*, NATO AGARD, Neuilly-sur Seine, France, 1990, pp. 3/1–3/17.

[12] Lee, K., Kerns, K., Bone, R., and Nickelson, M., "Development And Validation Of The Controller Acceptance Rating Scale (CARS): Results Of Empirical Research," *4th USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, New Mexico, USA, 2001.

[13] Endsley, M.R., Bolte, B., and Jones, D., "Automation and Situation Awareness," *Designing for Situation Awareness: An Approach to User-Centered Design*, CRC press, Boca Raton, Florida, USA, 2003, pp. 173–191.

[14] 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, California, USA, 2016, pp. 1–14.

[15] *737-6Q8/-7Q8/-8Q8 Flight Crew Operations Manual*, The Boeing Company, Seattle, Washington, USA, 2005.

[16] Kramer, L.J., Etherington, T.J., Last, M.C., Kennedy, K.D., and Bailey, R.E., "Quantifying pilot contribution to flight safety during drive shaft failure," *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, IEEE, St. Petersburg, Florida, USA, 2017, pp. 1–7.