Cognition, Technology & Work (2021) 23:353–365
https://doi.org/10.1007/s10111-020-00629-w

ORIGINAL ARTICLE

Designing flight deck applications: combining insight from end-users and ergonomists

Katie J. Parnell[1] · Victoria A. Banks[1] · Craig K. Allison[2] · Katherine L. Plant[1] · Peter Beecroft[3] · Neville A. Stanton[1]

Received: 27 November 2019 / Accepted: 16 March 2020 / Published online: 26 March 2020

Abstract
Technological advancement brings opportunities for enhanced information, support, and functionality within the flight deck. Whilst this has many benefits to the pilot and the overall safety of the aircraft, the practical integration of new technologies needs to be carefully considered throughout the entirety of the design process. The application of Human Factors methods must ensure that new technologies do not expose the system to new failures. This paper compares two methods of generating design recommendations for new technological features; the system human error reduction and prediction approach (SHERPA) and the Design with Intent (DwI) method. The assimilation of the recommendations from both methods presents interesting findings that highlight the benefits of integrating end-users within structured Human Factors methods to generate effective and usable technological interfaces. Case examples showing the similarities and differences between the concepts that the two methods generate are presented. The practicalities in using each approach within a Human Factors-driven design process are also discussed. The findings highlight the importance of end-user engagement in the early phases of the design lifecycle and how this relates to a Human Factors approach to design.

Keywords SHERPA · Design with intent · Interface design · Human error identification

1 Introduction

The potential for advanced technological features and applications to be integrated into the cockpit of large aircraft involved in commercial air transport offers many benefits to airlines, pilots, and the maintenance of safety for all in the aviation community (Salas et al. 2010). The cockpit of larger commercial aircraft currently provides pilots with extensive information regarding the current status of the aircraft as well as context relevant procedures. Yet, with the advancement of technology, enhanced and increasingly precise information can be delivered to pilots to inform them of aircraft status in a timely and assistive manner. One example is an engine condition monitoring tool that can inform pilots of issues residing within an aircraft engine in a timelier manner than can be currently achieved. This will allow issues to be more readily resolved, promote preservation of the engines, and enhance operational procedures.

A specific example which will be focused on within this paper is the opportunity for enhanced information relating to a possible oil leak within the aircraft engine. In current larger commercial aircraft, the pilot is only alerted to low oil pressure in the engine once the oil levels have reached critical levels. At such a point, they must then take immediate action to limit any adverse consequences and maintain flight safety. As oil levels have already reached low levels when this information is received, pilots are under time pressure to undertake corrective actions. Whilst information relating to oil level is currently available, pilots have to

Relevance to human factors/relevance to ergonomics theory: Human error identification (HEI) methods including System Human Error Reduction and Prediction Approach have been demonstrated to be of great value in improving the safety of systems. Such methods can be criticised, however, for being overly conservative and lacking insight from end-users. The current paper presents an initial comparison between HEI and designs generated by end-users in a series of design workshops. Results suggest that both methods are of value when developing novel systems.

✉ Craig K. Allison
Craig.Allison@solent.ac.uk

[1] Transportation Research Group, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, UK

[2] School of Sport Health and Social Sciences, Solent University, East Park Terrace, Southampton SO14 0YN, UK

[3] Rolls-Royce PLC, Derby, UK

actively search for this information which is buried within menu systems. Pilots would have to be actively monitoring this page during flight to determine if there was an oil leak. As an oil leak is a rare occurrence, pilots do not check oil levels frequently during flight. An advanced system could, however, be implemented to accurately monitor engine oil levels and inform the pilot of a suspected leak in advance of critical levels being reached. Such a system could provide enhanced decision support at appropriate moments, providing pilots more time to make a safe decisions and minimise operational disruption.

Despite these potential benefits, it is important to fully understand how such a tool could be integrated into the flight deck alongside the other functions and procedures that pilots must conduct. When making improvements to current systems to reduce the possibility for error, it is also important not to introduce new opportunities for misuse (Kirwan 1998a, b; Stanton et al. 2009).

With this in mind, the certification of new technologies into the flight deck requires the implementation of a comprehensive Human Factors assessment relating to the design of the device and its integration with other features already on the flight deck. Although Human Factors can, and should, be used throughout a system development (Sætren et al. 2016), Human Error Identification (HEI) methods are particularly valuable early within the design cycle to determine potential design-induced errors that may occur within human–machine interaction (Baber and Stanton 1994; Stanton et al. 2006, 2013). The findings of such analysis can then be used to provide measures with which to remedy the errors and inform design. The System Human Error Reduction and Prediction Approach (SHERPA, Embrey 1986) is an HEI method that aims to analyse system performance and identify errors induced by human operators and/or the design of a system. To adequately address this need, Hierarchical Task Analysis (HTA; Annett et al. 1971; Shepherd 2001; Stanton 2006) is combined with the SHERPA taxonomy of external error modes (Embrey 1986). This allows for a task analysis of the users' interaction with a system and the identification of potential error failures within each task. SHERPA is predicated on the assumption that errors are predictable (Embrey 1986). SHERPA has been applied in the aviation domain, informing design and flight safety (Harris et al. 2005), and has been advocated for its ease and reliability of application (Stanton et al. 2002). HEI methods are often critiqued, however, for their lack of validity, due to the requirement of subjective judgement on the analyst, as different analysts are likely to generate different conclusions (Stanton et al. 2013). Despite this, SHERPA has been found to be the most promising HEI method (Kirwan 1992), with encouraging validity and reliability findings (e.g., Baber and Stanton 1996; Stanton and Stevenage 1998; Hughes et al. 2015).

Applying SHERPA to the current tools used to assess engine condition monitoring can identify potential errors in the current system and also generate recommendations for remedial measures to overcome the identified errors. This process can, however, omit any other design concepts that may suggest radical design changes if they do not immediately relate to errors in the current system. Early stages of the design process, where SHERPA is utilised, is also an essential stage at which the views of the end-user of the system can be incorporated to ensure that future designs are usable, safe, and that they are well integrated with the other features of a system (Gould and Lewis 1985). The inclusion of potential end-users of the system within the early design phases is a valuable way of ensuring that designs reflect the users' requirements (Kujala 2003, 2008).

In the field of aviation where the user is a highly trained professional, their integration into the design process can pose additional complexities. Availability and engagement of users can be difficult to achieve, yet the benefits of pilot integration into the design process are essential to cater to their knowledge base. Furthermore, they are best suited to understand how possible designs may influence their interactions with the other aspects of the flight deck and, therefore, the viability of a design.

1.1 User-centred design

Norman and Draper (1986) stated that "to understand successful design requires understanding of the technology, the person, and their mutual interaction…" (p1). User-centred design places the user at the centre of the design process and enforces that the equipment is designed to meet the skills, knowledge, and abilities of the target user of the device (Harris 2007). This is key within the early stages of the design process (Cacciabue, and Martinetto 2006), rather than being considered when it may be too late to make adequate design changes (Gould and Lewis 1985).

Regardless of domain (Hesse et al. 2011; Frison et al. 2017), the inclusion of potential end-users within the design process is essential when proposed future designs includes changing the role and tasks of operators (Kaber et al. 2002). Automating specific functionality or changing the information that is presented requires users to adjust their interactions with a system or work environment (Kaber et al. 2002; Parasuraman and Wickens 2008). Furthermore, proposed novel designs do not always consider how such systems may be adopted by end-users. It is, therefore, important to understand how the outputs of an HEI method, such as the SHERPA, relate to the end-users' perspective of the current system and the generation of users' own design requirements. User-centred design acts as a bridge to achieving this (Karat 1997; Gould and Lewis 1985; Bekker and Long 2000; Kujala 2003). To ensure that

this occurs effectively, users' inputs must be included at the earliest opportunity, ideally from device conception, and throughout the design, testing, and integration stages of product development (Gould and Lewis 1985; Kujala 2003, 2008). One method that allows for the inclusion of the user in the design process is the Design with Intent method (DwI; Lockton et al. 2010), which allows for novel interface and system ideas to be generated (Read et al. 2016; Allison and Stanton 2020a, b).

1.2 Design with Intent (DwI)

The DwI method (Lockton et al. 2010) was developed to support novel product refinement, with a focus on user-centred design. The approach aims to generate novel design concepts by engaging future users of the system in idea generation sessions. The method focuses around a set of 101 design cards that act as suggestion aids to prompt discussion and generate novel ideas. Whilst DwI remains an underutilised methodology, Read et al. (2015) used the method in an integrated Cognitive Work Analysis Design Toolkit (CWA-DT), which applied to a design case study of a public transport ticketing system. Allison and Stanton (2020a, b) have also recently demonstrated the use of DwI in facilitating a creative process for generating design concepts for interfaces to promote fuel-efficient driving behaviour in motor vehicles.

As identified, insights gathered from HEI methods, including SHERPA, offer potentially useful, yet conservative improvements to a system. DwI, in contrast, offers the potential for radical redesign of a system, driven by end users' needs and requirements. Yet, it is unclear the extent to which the insights offered by SHERPA and DwI are comparable, complimentary, or potentially contradictory. The focus of this paper will, therefore, be to compare end-user interface ideas generated from use of the DwI method, with remedial design measures generated using SHERPA, within the context of flight deck technology. Reviewing the SHERPA and DwI outputs in parallel will identify how effective they are individually, noting similarities and differences as well as determining what can be gained by reviewing findings across methods.

2 Method

The methodology used to generate remedial design measures in the SHERPA and the design idea generation in the DwI workshops is detailed below. The method of mapping the reports from the pilots in the DwI workshops to the remedial measures identified in the SHERPA is then presented.

2.1 SHERPA

SHERPA was applied to determine the possibility for error occurring during current practise in responding to a suspected engine oil leak in the aircraft and propose new design ideas that could prevent these errors. To complete this task, an initial Hierarchical Task Analysis (HTA; Annett et al. 1971; Stanton 2006) was developed. Interviews were conducted with current airline pilots of larger commercial aircraft to inform the development of this HTA and identify the tasks involved in both detecting and managing the oil leak. Once the tasks required to manage the oil leak were understood, the possibility for error to occur within each of these tasks was assessed. Tasks from the bottom of the HTA (which represent the specific individual tasks that comprise all of the higher level goals of the process) were then reviewed for all possible errors that could arise across each stage in the process and within the system (Parnell et al. 2019).

2.1.1 Participants

Six pilots (two female and four male) with an Airline Transport Pilot License (ATPL) or Commercial Pilot License (CPL) for fixed-wing aircraft were interviewed. This was the point at which data saturation was reached (Grady 1998; Saunders et al. 2018). It was, therefore, deemed the cut off for the number of participants required for the analysis. Participants ranged in age from 26–35 years ($M = 30.17$, S.D. = 3.02). Participants had an average of 3692 h of flight experience (range = 2900–4500, S.D. = 635) and 8.08 years of experience since obtaining their pilots license (range = 5.5–10 years, S.D. = 1.74). Each pilot was reimbursed for their time spent conducting the study and any travel expenses incurred. The interviews were run in accordance with the University of Southampton Ethical and Research Governance Office policies (ERGO ID: 40619).

2.1.2 Procedure

Pilots were interviewed individually to obtain information relating to their response to a suspected oil leak in a current aircraft system. This utilised the Schema Action World Research Methodology (SWARM; Plant and Stanton 2016), which was developed to obtain information from pilots surrounding their decision-making processes and can be applied to understand what actions are available to pilots. Interviews were audio recorded and the transcripts were used to inform the development of the HTA. The main goal that was the given starting point for the HTA was to 'manage a suspected oil leak'. A total of 78 tasks were identified in the HTA. Once the HTA was completed, it was reviewed by both a Human Factors expert with over 30 years of experience

and an experienced pilot with over 10 year flight experience (see Parnell et al. 2019 for further details of the HTA). This ensured that all relevant information was captured. The bottom-level tasks from the HTA were then reviewed in the SHERPA.

The SHERPA error taxonomy was used to determine the possible errors that occur within each of these low-level tasks in the system and, therefore, manifest into errors within the wider system. The taxonomy classifies errors as one of the following: action, checking, retrieval, communication, or selection errors. There are different error types within each of these; for example, checking errors include: 'check omitted', 'check incomplete', 'check mistimed', etc. Reviewing the tasks with the error taxonomy and determining which errors may feasibly occur determined all possible errors. For each of the potential errors identified, remedial measures were proposed using improved design. A current commercial airline pilot with over 10 years of flight experience then reviewed the SHERPA for the errors identified and the remedial measures suggested, to determine if they were viable and appropriate.

2.2 Design with Intent

The Design with Intent (DwI) method that was used in this study followed the prescriptive model outlined by Lockton et al. (2010), whereby a predetermined set of the design cards were selected by the researchers as being relevant to the interface under assessment and the target behaviours identified in the HTA. The study was approved by the University of Southampton Ethical and Research Governance Office (ERGO: 41697).

2.2.1 Participants

A total of five participants were recruited across three workshops (two male and three female), aged between 31 and 38 years ($M = 34.6$). All participants were qualified airline pilots with a fixed-wing ATPL or CPL qualification held for a range of 9–10 years ($M = 9.7$ years). They had an average of 4140 h of flight experience, ranging from 3000 to 5000 h. All were currently employed by a commercial airline.

2.2.2 Equipment

Information about the study and the system to be designed were presented to participants on a PowerPoint presentation, the DwI design cards were then sequentially presented to participants. Participants were encouraged to draw their design ideas on large sheets of paper with coloured pens. The workshops were audio recorded to capture the discussions for the researchers to reference when they came to reviewing the designs.

2.2.3 Procedure

At the start of the workshop, the participants were presented with the following scenario:

During normal operational flight, you are alerted to a suspected oil leak following a warning signal on the flight deck. You must determine the criticality of the oil leak and take appropriate action.

Following presentation of this scenario, participants were asked to design a flight deck interface that would be of value when faced with such a challenge. Participants were asked to draw an initial design idea, which would act as an initial concept that could be amended as the session progressed. Participants were given the option to develop an independent design or produce a collaborative design, although pilots in all sessions chose to collaborate on their design. Once this initial design concept was completed, the researchers presented the DwI cards. A down-selected sample of 40 DwI cards was used (see Table 1). The researcher provided a summary of the meaning of each card upon presentation for clarity.

Table 1 Selected design cards from the each of the design lenses that were selected for use in the design workshop with pilots

Architectural lens	*Perceptual lens*
Angles	Colour associations
Conveyor belt	Contrast
Feature deletion	Implied sequences
Hiding things	Nakedness
Pave the cow paths	Perceived affordances
Positioning	Prominence
Segmentation and spacing	Proximity and grouping
Simplicity	Similarity
	Transparency
Interaction lens	*Machiavellion lens*
Karios	Anchoring
Partial completion	Forced dichotomy
Peer feedback	Format lock in/out
Progress bar	Slow/no response
Real time feedback	
Simulation and feedforward	
Summary feedback	
Tunnelling and wizards	
Error proofing lens	*Cognitive lens*
Are you sure?	Commitment and consistency
Choice editing	Do as you're told
Interlock	Scarcity
Opt-outs	
Portions	
Task lock in/our	
Ludic lens	*Security lens*
Levels	Where are you?

Following the presentation of each card, participants discussed if they felt that the suggestion on the card was relevant to the scenario and whether it could be incorporated into their design. This process was repeated for each card until all 40 down-selected design cards had been discussed. Once this process was completed, participants were asked to draw a final design concept incorporating what they had discussed and the changes which they had made to their initial concept.

2.3 Mapping DwI workshop responses to SHERPA errors and remedial measures

As noted previously, the completed SHERPA generated a complete list of possible errors that could occur when pilots are managing a suspected oil leak in the current system, as well as a complimentary list of remedial measures through which design could address these errors. Conversely, the DwI workshops generated a number of design concepts from potential end-users. The outputs from these methods were compared to determine whether the remedial measures suggested in each were related, or conversely presented opposing ideas. It should be noted that it was not the intention of the DwI workshops to generate a list of currently possible errors, so this was not the focus of comparison. Rather, the focus of the comparison was the extent to which the design concepts generated by end-users addressed errors identified within the current system.

A table was constructed listing the errors identified using SHERPA, alongside their corresponding remedial measures. The generated design concepts from the DwI workshops were then reviewed against this list. The review determined if the design concepts matched the remedial measures developed using SHERPA, conflicted with the remedial measures developed using SHERPA or were novel ideas that had not been identified using SHERPA. Table 2 presents the inputs, equipment, and the users required and the expected outputs for the SHERPA analysis and DwI workshops. This highlights that while the SHERPA utilises input from expert users to inform the development of the HTA, independent researchers predominantly drive the analysis. DwI, in contrast, is developed and driven by the insight of potential end-users.

3 Results

3.1 SHERPA

The SHERPA analysis identified a total of 108 potential errors when responding to a suspected engine oil leak, the most frequent error type being action errors ($n = 36$) that related to omitted actions or conducting the wrong actions. The most frequent specific error type, however, was obtaining wrong information ($n = 19$), which was a retrieval failure. There were multiple errors that occurred frequently that could be tackled using the same remedial design measures. These errors were aggregated to identify 19 key errors, and their corresponding key remedial measures (see Table 3).

It is notable that while there may appear to be a surmountable number of errors that relate to the pilots' actions, or lack thereof, there is considerable responsibility on the interface designers to encourage usable and effective interfaces that limit the opportunity for failures to occur in the system. For example, while it could be suggested that a pilots' failure to notice a warning signal (error 5) is pilot error, if the warning signal was of greater salience and placed in a location relevant to the anticipated task being conducted and then pilots would be in a better position to notice it.

3.2 DwI workshops

The three DwI workshops each generated a different design concept for a future engine health-monitoring system that could be used to better inform the pilot of the engine status and a suspected oil leak. These findings highlight the key ideas that the pilots noted in their discussions and prioritised

Table 2 Requirements of the SHERPA and DwI workshops

Method	Input	Equipment	Users	Outputs
SHERPA	Pilot interviews HTA	Pen/paper (or computer design programme) for the SHERPA	3× Human Factors researchers 1× Pilot SME	Predicted errors in the current system Remedial measures to overcome current system predicted errors
DwI workshops	Oil leak scenario	DwI cards Pen/paper for drawing	2× Human Factors researchers 5x Pilot SMEs	Novel design concepts Discussion on each DwI card relating to design requirements

Table 3 Key errors identified from the SHERPA and their proposed remedial measures

Error failure	Remedial measure
Failure to check oil page occasionally as part of SOP	Prompts to check oil on computerised checklist that direct the pilot; increase saliency of the display
Fail to read oil level/temperature/pressure correctly	Visual representation of oil level/temperature/pressure above threshold level (bar display)
Fail to compare oil level/temperature/pressure to correct limits	Bring up current oil level/temperature/pressure (bar display) with historic and predicted oil levels (graph display)
Fail to notice warning signal	Make warning more salient, escalation of warning
Fail to recognise warning as oil leak and check correct information	Highlight area associated with the warning
Fail to correctly trend oil temperature/pressure	Automatically trend current and predicted oil pressure on graphical display
Fail to compare oil parameters to limits correctly	Automatically trend current and predicted oil on graphical display including a 'limit' line
Fail to use other information sources to check for oil leak	Prompts to check evidence of oil leak external to the aircraft on computerised checklist,
Fail to check the correct engine	Check list directs to the side the engine oil leak is reported on
Fail to use other information sources to check for oil leak	Prompts to contact the cabin crew and check for evidence of oil leak external to the aircraft on computerised checklist
Fail to move the throttle leaver correctly	Display indication to show leaver and level/direction to move it in
Fail to adjust calculation of trend for remaining flight	Automatically trend current and predicted oil parameters in response to updated flight parameters (e.g. throttle)
Fail to correctly check distance to alternative destinations	Highlight area associated with alternative destinations on FMC
Fail to check emergency facilities/maintenance facilities/weather correctly	Include up-to-date maintenance facilities/emergency facilities/weather on FMC/nav display and highlight area associated
Fail to check runway length correctly	Include up-to-date emergency facilities on FMC/nav display and highlight area associated
	Highlight area associated with runway length on FMC
Fail to switch engine off at correct time	Add to computerised checklist
Fail to contact Air traffic control	Auto dial air traffic control
Fail to correctly choose to hide warning signal	Highlight area associated with the warning

in their design. An example of the ideas generated and how they relate to the cards is given in Table 4.

A complete discussion of the generated design concepts from the DwI workshops is not the main focus of this paper, and as such, these will not be not discussed at length. Instead, the focus of this paper is to assess how the design concepts generated by the pilots in the DwI workshops compared to the design recommendations generated by Human Factors researchers using SHERPA.

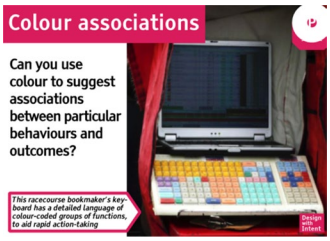
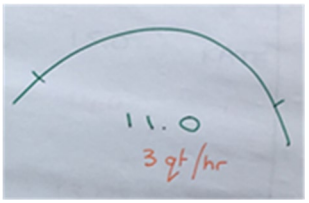
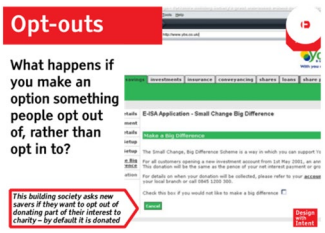
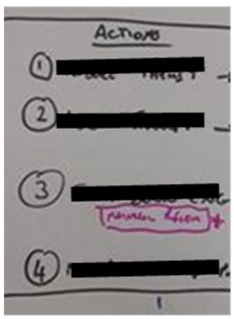
3.3 Comparing SHERPA and DwI workshops

Within design concepts, there were several pertinent ideas that consistently emerged; however, there were also some conflicting ideas. The level of detail and insight that could be gleaned from involving pilots in the design process was evident, as well as their generation of solutions to the problems which they identified from their perspective. An example excerpt from the table is presented in Table 5.

Using the complete table contrasting the DwI recommendations to the SHERPA errors and recommendations, it was possible to identify which ideas from the DwI were a match with the SHERPA, conflicted with the SHERPA or were new ideas outside of the SHERPA. The frequencies of each of these discrete occurrences were calculated and are presented in Table 6. This shows the 'Matches' which are the recommendations that are also generated in the SHERPA method and the 'Conflicts' which are ideas that were not generated in the SHERPA method. From Table 6, it can, therefore, be seen that there were 29 incidences of matched recommendations. There were nine cases where the SHERPA ideas conflicted with the recommendations made by pilots in the DwI workshops. There were 96 recommendations made in the DwI workshops that were not reported in the SHERPA. The ∞ represents all other ideas that neither method generated, the potential of which cannot be accounted for in this analysis.

Table 6 gives an indication that there was some overlap in the design recommendations that the two methods proposed and thus demonstrates that using SME participants to generate ideas independently of HEI methods can validate the outputs. Yet, there were also recommendations

Table 4 DwI cards and the example design recommendations that they generated

Lens	Card	Design recommendation	Illustration
Perceptual		Use green/amber/red for consistency. Code numbers on the oil gauge that show rate of oil loss	
Interaction		Provide a count-down timer to show how much time there is until oil starvation in the engine	
Error proofing		Give a list of alternative airports in case of a diversion, provide recommended actions but allow the pilot to opt out of this and make their own selection	

that each method proposed measures that were not supported by the other. A key area of concern here is the nine recommendations made using SHERPA that conflict with the SME recommendations made in the DwI design workshops. This suggests that the SHERPA alone may not be an ideal method of capturing design measures that are practical and useful to the user of the system. That is to say, we should not rely solely on analyses conducted without the involvement of representative end-users. As an error identification system, the clear lack of identification suggests a weakness in the method, or the need for the method to be suitably validated by SMEs. It is also interesting to note the 96 new design ideas that could resolve the errors identified using SHERPA that were generated from DwI discussions, but were not identified using SHERPA. This demonstrates the rich data generated by the DwI workshops, and further supports the value of end-user input into design generation.

To illustrate the impact of the differences and similarities across the two methods, three case examples are presented. These aim to show how the SHERPA presented measures to minimise errors that are in part supported by the users of the system, yet accounting for the users input allows for a broader picture in how new technologies can be integrated into the flight decks

3.4 Case examples

3.4.1 Case 1

Automatically trend oil levels for the remaining time of flight.

SHERPA error: Fail to adjust calculation of oil temperature/pressure leak trend for the remainder of flight.

SHERPA remedial measure: Automatically trend current and predicted oil parameters in response to updated flight parameters.

3.4.2 DwI similarities

Pilots identified the need to update the predicted oil pressure/temperature levels in line with the flight parameters. It was suggested that the option of having access to information related to the trend of the oil over different time periods and in relation to different parameters depending on the status of the flight would be beneficial. They suggested that

Table 5 Extract from the table contrasting the DwI recommendations to the SHERPA errors and recommendations

SHERPA Error	SHERPA remedial measures	DwI		
		Match	Conflict	New ideas
Fail to compare oil level/temperature/pressure to correct limits	Bring up current oil level/temperature/pressure (bar display) with historic and predicted oil levels (graph display)	"i button" only appears when needed and then it gives information on the next phase of flight that is required Location of the oil level on the dial needs to be easy to read	Pilots need freedom to look at what they want	If something more important happens, it needs to be able to be overridden (task lock in/out—pop-up checklists with actions) Different pages of information and not presenting too much information on one page to prevent clutter
Fail to adjust calculation of trend for remaining flight	Automatically trend current and predicted oil parameters in response to updated flight parameters (e.g., change in throttle)	Option of reviewing the predicted trends across different time periods, e.g., 10 min or 30 min. This would give different rates that may be required for different types of flights Real time feedback could be given based on their action	Too complicated to simulate and understand every possible option. Keep it simple	Count down timer on how much time you have left under oil starvation rate number could be added to the dial on primary screen

Table 6 Frequency of ideas generated in the DwI that were present or absent in the SHERPA and vice versa

SHERPA	DwI	
	Match	Conflict
Match	29 Proposed in both the SHERPA and DwI	9 Proposed in the SHERPA but not the DwI
Conflict	96 Proposed in the DwI but not the SHERPA	Not proposed in DwI or SHERPA

real-time feedback on the oil leak status could be given in response to the actions that they carried out in their attempt to manage the situation, for example, reducing the throttle and powering down the engine.

3.4.3 DwI differences

Divergent ideas also emerged, with some pilots, suggesting that the option to simulate every possible option and possibility would be too complicated to process and understand in a scenario such as this. Pilots cautioned the proposition of overly complex information and the presentation of the multiple possible actions which they could take, favouring instead easy to understand, and real-time feedback on actions which they had taken.

3.4.4 DwI new ideas

Pilots promoted the need for information regarding the 'time until oil starvation' and were, therefore, keen to have this information presented to them in a clear and easily accessible manner. While access to detailed information on the oil leak trend on secondary displays was suggested, some pilots also suggested bringing information on the rate of change or the 'time until oil starvation' to the primary display when there was a suspected oil leak, as described in this scenario. This could involve a count-down timer on the main display or an indicator of the rate of oil leak on the primary oil level display. Noting that if there was no oil leak, this information would not be needed and, therefore, should not appear, on the primary display.

3.4.5 Case 2

Provide up-to-date contextual information on landing options and parameters that may influence the decision to continue with the flight.

SHERPA error: Fail to check emergency facilities/maintenance facilities/weather correctly.

SHERPA remedial measure: Include up-to-date maintenance facilities/emergency facilities/weather on FMC/navigation display and highlight area associated.

3.4.6 Dwl similarities

Pilots unanimously felt that it was important to have up-to-date information regarding their options for continuing or diverting the flight. Key factors in this decision included weather, terrain, location, and information relating to runway length. They were also keen to have information about the emergency facilities at possible divergence locations as well as maintenance for the aircraft to minimise overall disruption. Primarily pilots' responses focused on the importance of maintaining the condition of the engine for the sake of the airline.

3.4.7 Dwl differences

Pilots were not keen to have systems that automated the decision-making process, nor did they want their decisions to be led by the type of information that was presented or the way in which it was presented. The pilot is given a great deal of training to undertake their job and they like to exercise this and be given the freedom and responsibility to make the decisions using all of the information that they are trained to do. Hence, leading the pilots' decision-making or removing freedom of the decision-making process in not well regarded. Designers must, therefore, treat the user with respect and enhance their engagement with it.

3.4.8 Dwl new ideas

The possibility of having a list of proposed options available given the state of the oil leak, the current weather conditions, location, status of flight, and ground facilities was consistently discussed by pilots. They proposed that these options would be preprogramed in relation to the route that they were to undertake and would be presented in case of such a scenario. Pilots noted that they may be interested to note the preferred diversion option of the company, which would presumably minimise airline incurred costs and overall disruption. Yet, they also wanted the ability to opt out of this option if they perceived there was a better solution. Pilots did not want to know the reasons why the company route was preferred as it may inadvertently influence the pilots' decision. Pilots also generated novel ideas regarding the ability to contact maintenance teams on the ground to get feedback on the status of the leak.

3.4.9 Case 3

Provide prompts to check oil using computerised checklist.

SHERPA error: Fail to spot oil leak due to not checking oil page, as included in the standard operating procedures (SOP).

SHERPA remedial measure: Provide prompts to check oil on a computerised checklist to regulate when to check it and prevent it from being forgotten.

3.4.10 Dwl similarities

The benefits of spotting the oil leak before the warning signal through standardised checking mean that the leak can be dealt with before it reaches dangerous levels or engine oil starvation. Pilots suggested that they value both the oil information currently available and the oil leak warning system that cues assessment of oil levels and investigation into the possibility of a leak. Some pilots suggested having a button linked to the Quick Reference Handbook (QRH) that informs pilots of advised oil levels for different pressures and temperatures to guide their checking process.

3.4.11 Dwl differences

There were conflicting views among pilots on the automatic nature of the prompts with some pilots' keen not to have prompts popping up to guide them. Not all pilots wish to be given extensive direction by the system as they have inherent knowledge of aircraft and engine operating parameters gained from training and inflight experience. The pilots still wanted to have the autonomy to decide when to look-up information and act in the way which they saw fit for the current situation.

3.4.12 Dwl new ideas

Pilots across all workshops were clear that their taught priorities within the flight deck are to 'Aviate, Navigate, and Communicate', in that order. Therefore, the need to check oil level can often become downgraded. While pilots do monitor oil level, it can often be difficult to identify small changes, as the detail is very small. This is why, they must rely on the oil warning system indicating when an oil leak is occurring, with the option then being given to review the trend in oil pressure/temperature levels. Yet, pilots suggested that they still wanted the option to be able to dip in and out of the display to focus on their priorities surrounding the need to 'Aviate, Navigate, and Communicate'. Pilots wanted to be prompted that the problem did exist and so suggested the ability to minimise the oil information screen to focus on immediate tasks and then automatically reopening the oil page once the other tasks were complete. In this way,

the pilots were keen to have a reconfigurable display with pages that could be moved and alternated in favour of the current priorities.

4 Discussion

This paper has focused on the development of a new system that has the potential to provide pilots with better information relating to the status of the engine in a case scenario of an engine oil leak. Two methods were applied that can assist in the development of the interface for integrated engine health-monitoring application on the flight deck; a traditional approach seeking remedial error reduction, led by researchers (SHERPA; Embrey 1986), and radical redesign approach led by end-users (DwI; Lockton et al. 2010).

Using SHERPA, 19 key errors within the current system for managing an oil leak scenario were identified and 19 corresponding design measures were proposed to remedy these errors. The recommendations made by pilots across three DwI workshops were noted, and design ideas that pilots generated within the design workshops were mapped to the errors/remedies identified using SHERPA. Similarities in the design ideas gleaned from the DwI workshops to remedial measures generated using SHERPA were apparent, supporting the previous work in validating the utility of SHERPA (Stanton et al. 2002). Yet, there were also incidences where suggestions from the DwI workshops directly conflicted with the remedial measures proposed using SHERPA. This suggests that caution is needed when using SHERPA to recognise the desires and capabilities of end-users. Furthermore, the DwI workshops generated a large number of design ideas and suggestions that were unforeseen using SHERPA, potentially highlighting the importance of using multiple approaches. This suggests that the insights gained from SHERPA and DwI workshops are complimentary, highlighting the need for input from both Human Factors practitioners in combinations with end-users of a system.

The use of HEI methods has previously been highly useful in capturing errors within current systems and providing recommendations to practise and design for future system iterations to minimise error potential (Stanton et al. 2002; Lane et al. 2006; Hughes et al. 2015). A key benefit of SHERPA is its ability to account for a broad scope of factors and implement these into the remedial measures. This has been effective in preventing errors from being continually attributed to the human (Lane et al. 2006). It is also an accessible methodology, with novices to a domain able to apply the approach to a similar standard as those with lengthy experience in the area (Stanton et al. 2002). In comparison to other HEI methods, such as the Human Error Template (HET; Marshall et al. 2003), SHERPA is praised for its ability to enable the generation of solutions alongside system failures. For such reasons, SHERPA has been suggested to be the best tool for analysing human error within aviation and enabling the errors to be designed against (Harris et al. 2005).

Despite the clear accessibility and advantages of SHERPA, it was also evident that end-users, in the current case pilots, can propose ideas that counter those made by Human Factors researchers using this approach. In addition, it was also evident that end-users are capable of generating a wealth of other ideas that may greatly assist the design process. These novel ideas can be invaluable to the design process, but are frequently ignored when the end-user of the system is not represented within the overall design process (Kujala 2003). Through mapping the reports that the pilots gave within the DwI workshops to the errors and remedial measures that the SHERPA approach identified, it was clear that SHERPA was effective in generating usable remedial measures. However, it is also clear that end-users would not have approved of some of the developed remedial measures. Notably, it was evident that pilots have an experts' insight into how a future system would be integrated alongside all the other features and tasks that must be completed, of which SHERPA analysts would not be not fully aware of.

Case 3 suggested the provision of prompts to check the oil levels were a logical recommendation to the error that the SHERPA identified. Yet, in reality, pilots were not keen to have automatic prompts that could distract them and interfere with the priority to 'Aviate, Navigate, and Communicate'. While they realised the severity of an oil leak scenario, pilots are also tasked with lots of other possible scenarios and factors that they must monitor which were not considered within the SHERPA analysis or the remedial measures generated. It was also made clear that the pilots did not want a proposed system to take over or dictate their options for them in the oil leak scenario. The responses which they gave focused on the benefits of having more technologies to assist with their response to the scenario, through the provision of real-time feedback on the rate of oil loss and of contextual factors relating to possible diversions including weather, runway length, and emergency facilities. Yet, they did not want the system to lead their decision or control what aspects of the flight deck which they looked at and when. While it may seem beneficial to a non-expert in the domain to provide easy directions and checklists for responses that reduce errors through limiting input, in reality, pilots undergo extensive training and are highly skilled experts that want to be in control of the flight and use their knowledge to deliver the best outcome which they can for the safety of the flight (Schutte 2017) as well as acting in the interests of the airline.

4.1 User led design

The disparity between those with domain knowledge and those with Human Factors knowledge in conducting error analysis is suggested by Stanton and Baber (2002). The judgement of the analyst plays a large role in the output of HEI method application, yet Stanton and Baber (2002) identified that novice users of a system could apply such methods to an acceptable standard with ease. This was deemed to be due to the structured nature of the SHERPA and Task Analysis for Error Identification (TAFEI) that enabled a structure for judgements to be made, without constraining them (Stanton and Baber 2002). The findings from the current work demonstrate that while the Human Factors researchers were able to adequately predict possible errors within the system and determine remedial measures that mitigated these errors, final concept generation is much improved with the addition of end-user input. The insights offered by potential end-users added rich detail of how to increase engagement with a future system as well as its integration within the flight deck. This is an important consideration that practitioners without user expertise should be aware of when applying Human Factors methods within a focused design approach.

Furthermore, advocates of user involvement in the design process have documented the increased levels of the users' acceptance of the system and user satisfaction (Kujala 2003) as well as the cost-effectiveness in streamlining the iterative prototyping and usability evaluation that new systems must undergo (Chatzoglou and Macaulay 1996). Yet, it is appreciated that the early involvement of the end-user is not always easy to achieve, and to be effective, it requires the application of developed methods and structured roles of both the system and the user (Kujala 2003). It is often the case that experienced users of a system are not experienced system designers, and therefore, bridging this gap may be difficult. However, it is evident that there is value added by pilot SMEs to future technologies reporting on the health of the engine has suggested many possible new options for interface design, as well as highlighting key facets at a pre-conceptual stage in the design process that must not be incorporated into later stages of the design. It is, therefore, apparent that a combination of insight from both Human Factors practitioners and end-users would add the most value to the development of novel systems (Fénix et al. 2008).

There are, however, complexities in the involvement of end-users within design. Acquiring a representative group of users is important to capture relevant characteristics of the user base. Different characteristics are important to consider. For instance, experience levels can have an influence over design requirements with less experienced users maybe requiring more guidance than experienced users. Variations in age, gender, and culture should be considered. Caution should also be held not to over rely on user input but to keep a focus on the usability of the design elements. It was evident that the SHERPA is a tool that is driven by Human Factors researchers with input from users, whilst the DWI method is user-driven method with input from Human Factors researchers to decide the design cards to be discussed and steer the workshop. The DWI method generated a rich set of data, much of which was out of the bounds of the SHERPA method. The refinement of these rich data to viable and feasible opportunities for enhanced design is an essential next step.

4.2 Limitations

Both the SHERPA and DWI methods focused on the development of design in isolation from the wider aircraft cockpit. Therefore, they do not consider how the designs may actually relate to real-world integration and different contexts of use. Making amendments to enhance performance in some areas of a design can have knock-on adverse effects to other areas of the system. For example, the SHERPA measure to "Make warning more salient, escalation of warning" may in turn eclipse other warnings or interrupt the flow of information. Therefore, such measures need to be carefully considered at the integration stage.

Furthermore, the implementation of any design within the aircraft must undergo substantial analyses and verification to assure that it meets the required standards and can be certified for use (De Florio 2016). A benefit of the DWI user-led method is that it encourages radical and alternative thinking to design problems which in turn generates novel design ideas. These designs are, however, still subject to the same standards and certification processes as current technologies. Furthermore, due to the rigour in the aviation domain, the pilots' responses to a scenario such as an oil leak are limited by the manufacturer's procedures. Pilots have a low level of autonomy. The integration of such concepts must, therefore, be carefully considered in order to maintain their benefits whilst ensuring safe and reliable aircraft design.

4.3 Future work

Future work will strive to develop the key feasible design ideas that were generated through both the SHERPA and DwI methods, staying true to end-user feedback. The next stage in the design process will involve the generation of design mock-ups for further user evaluations. The use of layout analysis and heuristic evaluations (e.g., Nielsen 1994) can be utilised early within the design cycle to assess the usability of interfaces, even at a stage where only rudimentary diagrams are available (Stanton et al. 2013). This enables user input in a functional capacity with data obtained before effort has been expended in generating sub-adequate interface designs. Evaluation between different design

concepts can also be determined at this stage in the design process.

5 Conclusion

When designing devices that provide an interface between the human and a technological system it is important that it is usable and effective in its purpose, and that it does not introduce the opportunity for error that can adversely affect the safety of the system. SHERPA has been used in the aviation domain to predict possible errors and provide design recommendations. This paper has shown how insights from end-users, gained within design workshops can add valuable insights when applying SHERPA and provide recommendations for error prevention. The insights that Human Factors researchers gain from the application of these methods, even when they are validated by a SME, omits substantial detail regarding the wider functionality of the system and the implications which it may have within this. Through the application of the DwI method, greater insights into the preferences of the pilot user of the system have been obtained that have highlighted where recommendations made through the SHERPA may be effective and conversely disruptive. For the generation of usable and error resistant interfaces, it is important that the Human Factors methodologies are true to the preferences of the user and the functionality of the wider system as a whole.

Acknowledgements This project was co-funded by Innovate UK, the UK's Innovation Agency, with support from the UK Aerospace Technology Institute. We would also like to thank the pilots who participated in the work contributing to this paper.

References

Allison CK, Stanton NA (2020a) Ideation using the "Design with Intent" toolkit: a case study applying a design toolkit to support creativity in developing vehicle interfaces for fuel-efficient driving. Appl Ergon 84:103026

Allison CK, Stanton NA (2020b) Constraining design: applying the insights of cognitive work analysis to the design of novel in-car interfaces to support eco-driving. Automot Innov 3:30–41

Annett J, Duncan KD, Stammers RB, Gray MJ (1971) Task analysis. Department of employment training information paper 6. HMSO, London

Baber C, Stanton NA (1994) Task analysis for error identification: a methodology for designing error-tolerant consumer products. Ergonomics 37(11):1923–1941

Baber C, Stanton NA (1996) Human error identification techniques applied to public technology: predictions compared with observed use. Appl Ergon 27(2):119–131

Bekker M, Long J (2000) User involvement in the design of human–computer interactions: some similarities and differences between design approaches. In: McDonald S, Waern Y, Cockton G (eds) People and computers XIV (Proceedings of HCI'2000). Springer, London, pp 135–147

Cacciabue PC, Martinetto M (2006) A user-centred approach for designing driving support systems: the case of collision avoidance. Cogn Technol Work 8(3):201–214

Chatzoglou PD, Macaulay LA (1996) Requirements capture and analysis: a survey of current practice. Requir Eng 1(2):75–87

De Florio F (2016) Airworthiness: an introduction to aircraft certification and operations. Butterworth-Heinemann, UK

Embrey DE (1986) SHERPA: a systematic human error reduction and prediction approach. In: Proceedings of the international topical meeting on advances in human factors in nuclear power systems. Knoxville, TE

Fénix J, Sagot JC, Valot C, Gomes S (2008) Operator centred design: example of a new driver aid system in the field of rail transport. Cogn Technol Work 10(1):53–60

Frison AK, Pfleging B, Riener A, Jeon MP, Alvarez I, Ju W (2017) Workshop on user-centered design for automated driving systems. In: Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications adjunct, September, ACM, pp 22–27

Gould JD, Lewis C (1985) Designing for usability: key principles and what designers think. Commun ACM 28(3):300–311

Grady MP (1998) Qualitative and action research: a practitioner handbook. Phi Delta Kappa International, USA

Harris D (2007) A human-centred design agenda for the development of single crew operated commercial aircraft. Aircr Eng Aerosp Technol 79(5):518–526

Harris D, Stanton NA, Marshall A, Young MS, Demagalski J, Salmon P (2005) Using SHERPA to predict design-induced error on the flight deck. Aerosp Sci Technol 9(6):525–532

Hesse T, Engström J, Johansson E, Varalda G, Brockmann M, Rambaldini A, Fricke N, Flemisch F, Koster F, Kanstrup L (2011) Towards user-centred development of integrated information, warning, and intervention strategies for multiple ADAS in the EU project interactive. In: Stephanidis (eds) Universal access in human computer interaction: context diversity. International conference on universal access in human–computer interaction, July, Springer, Berlin, Heidelberg, pp 280–289

Hughes CM, Baber C, Bienkiewicz M, Worthington A, Hazell A, Hermsdörfer J (2015) The application of SHERPA (Systematic Human Error Reduction and Prediction Approach) in the development of compensatory cognitive rehabilitation strategies for stroke patients with left and right brain damage. Ergonomics 58(1):75–95

Kaber DB, Riley JM, Tan K (2002) Improved usability of aviation automation through direct manipulation and graphical user interface design. Int J Aviat Psychol 12(2):153–178. https://doi.org/10.1207/S15327108IJAP1202_4

Karat J (1997) Evolving the scope of user-centered design. Commun ACM 40(7):33–38

Kirwan B (1992) Human error identification in human reliability assessment. Part 2: detailed comparison of techniques. Appl Ergon 23(6):371–381

Kirwan B (1998a) Human error identification techniques for risk assessment of high risk systems—Part 1: review and evaluation of techniques. Appl Ergon 29(3):157–177
Kirwan B (1998b) Human error identification techniques for risk assessment of high risk systems—Part 2: towards a framework approach. Appl Ergon 29(5):299–318
Kujala S (2003) User involvement: a review of the benefits and challenges. Behav Inf Technol 22(1):1–16
Kujala S (2008) Effective user involvement in product development by improving the analysis of user needs. Behav Inf Technol 27(6):457–473. https://doi.org/10.1080/01449290601111051
Lane R, Stanton NA, Harrison D (2006) Applying hierarchical task analysis to medication administration errors. Appl Ergon 37(5):669–679
Lockton D, Harrison D, Stanton NA (2010) The design with intent method: a design tool for influencing user behaviour. Appl Ergon 41(3):382–392
Mandal S, Singh K, Behera RK, Sahu SK, Raj N, Maiti J (2015) Human error identification and risk prioritization in overhead crane operations using HTA, SHERPA and fuzzy VIKOR method. Expert Syst Appl 42(20):7195–7206
Nielsen J (1994) Usability engineering. Academic Press, Boston
Norman DA, Draper SW (1986) User centered system engineering. Erlbaum, Hillsdale
Parasuraman R, Wickens CD (2008) Humans: still vital after all these years of automation. Hum Factors 50(3):511–520
Parnell KJ, Banks VA, Plant KL, Griffin TG, Beecroft P, Stanton NA (2019) Predicting design-induced error on the flight deck: an aircraft engine oil leak scenario. Hum Factors. https://doi.org/10.1177/0018720819872900
Plant KL, Stanton NA (2016) The development of the Schema World Action Research Method (SWARM) for the elicitation of perceptual cycle data. Theor Issues Ergon Sci 17(4):376–401
Read GJ, Salmon PM, Lenné MG (2016) When paradigms collide at the road rail interface: evaluation of a sociotechnical systems theory design toolkit for cognitive work analysis. Ergonomics 59(9):1135–1157
Sætren GB, Hogenboom S, Laumann K (2016) A study of a technological development process: human factors—the forgotten factors? Cogn Technol Work 18(3):595–611
Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, Burroughs H, Jinks C (2018) Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 52(4):1893–1907
Schutte PC (2017) How to make the most of your human: design considerations for human–machine interactions. Cogn Technol Work 19(2–3):233–249
Shepherd A (2001) Hierarchical task analysis. Taylor and Francis, London
Stanton NA (2006) Hierarchical task analysis: developments, applications, and extensions. Appl Ergon 37(1):55–79
Stanton NA, Stevenage SV (1998) Learning to predict human error: issues of acceptability, reliability and validity. Ergonomics 41(11):1737–1756
Stanton NA, Harris D, Salmon PM, Demagalski JM, Marshall A, Young MS, Dekker SWA, Waldmann T (2006) Predicting design induced pilot error using HET (Human Error Template)—a new formal human error identification method for flight decks. Aeronaut J 110(1104):107–115
Stanton NA, Salmon P, Harris D, Marshall A, Demagalski J, Young MS, Waldmann T, Dekker S (2009) Predicting pilot error: testing a new methodology and a multi-methods and analysts approach. Appl Ergon 40(3):464–471
Stanton NA, Salmon PM, Rafferty LA, Walker GH, Baber C, Jenkins DP (2013) Human factors methods: a practical guide for engineering and design. CRC Press, USA