ELECTRONIC CHECKLISTS: IMPLICATIONS FOR DECISION MAKING

Kathleen L. Mosier
Stanford University
Stanford, CA

Everett A. Palmer
NASA Ames Research Center
Moffett Field, CA

Asaf Degani
San Jose State University Foundation
San Jose, CA

Checklists are a way of life on the flight deck, and, undoubtedly, are indispensable decision aids due to the volume of technical knowledge that must be readily accessible. The improper use of checklists, however, has been cited as a factor in several recent aircraft accidents (National Transportation Safety Board, 1988, 1989, 1990). Solutions to checklist problems, including the creation of electronic checklist systems which keep track of skipped items, may solve some problems but create others. In this paper, results from a simulation involving an engine shutdown are presented, and implications of the electronic checklist and "memory" checklist are discussed, in terms of potential errors and effects on decision making. Performance using two types of electronic checklist systems is compared with performance using the traditional paper checklist. Additionally, a "performing from memory" condition is compared with a "performing from the checklist" condition. Results suggest that making checklist procedures more automatic, either by asking crews to accomplish steps from memory, or by checklists that encourage crews to rely on system state *as indicated by the checklist*, rather than as indicated by the system itself, will discourage information gathering, and may lead to dangerous operational errors.

INTRODUCTION

Checklists are a way of life on the aircraft flight deck, and, undoubtedly, are indispensable decision aids in an environment in which the volume of technical knowledge is enormous. They are designed to function as information safeguards and protectors against faulty memory retrieval. All of the high workload phases of flights (e.g., taxi, take-off, descent and landing, emergencies) are managed via checklists to ensure that no steps or procedures are forgotten. The improper use of checklists, however, has been cited as a factor in recent aircraft incidents and accidents (National Transportation Safety Board, 1988, 1989, 1990).

Degani and Wiener (1990) conducted a field study to analyze factors affecting the successful and unsuccessful use of air carrier checklists. They found that the standard paper checklists have a number of weaknesses, including the lack of a pointer to the current checklist item, the inability to mark skipped items, and the possibility of getting lost while switching between checklists. To address these weaknesses, two electronic checklist systems have been implemented in the Advanced Concepts Flight Simulator (ACFS) at NASA-Ames Research Center. An additional weakness of the standard paper checklist is the amount of time consumed in "getting out the checklist." To ensure quick response in emergency situations, many airline checklists for emergency situations, such as engine fires or hydraulic failures, have long included several Immediate Action Items. These are steps which are to be performed automatically from memory, before taking time to locate the pertinent checklist.

These "solutions" may solve some problems at the cost of exacerbating others. For example, reducing the time to perform "immediate action items," may also reduce the amount of conscious thought or analysis preceding their performance. In the recent crash of a British Midlands 737-400, the flight crew performed, from memory, the ENGINE FAILURE checklist, and shut down a healthy engine following the failure of the other engine (Department of Transport, 1990). With respect to the electronic checklist, although it eliminates some of the problems described above, it may introduce new errors by virtue of its automaticity and the fact that crews rely on the checklist as an indicator of system state rather than as a procedural aid. Palmer and Degani (1991), in fact, found that crews were less likely to detect system anomalies if the electronic checklist showed the items as completed.

The purpose of this paper is to discuss some of the implications of the electronic checklist and of "from memory" checklist procedures, and to present some preliminary data suggesting potential negative effects on crew decision-making processes.

Checklist Procedures and the Electronic Checklist

In a study designed to investigate the effectiveness of the electronic checklist for commercial air transport, 12 two-person, glass-cockpit crews flew a full-mission simulation in the NASA-Ames ACFS. Pilots were randomly assigned to one of the three checklist conditions: 1) automatic-sensed checklist; 2) manual-sensed checklist; or 3) paper checklist. The electronic checklists could be displayed on the lower portion of the captain's or first officer's systems display. Each of the

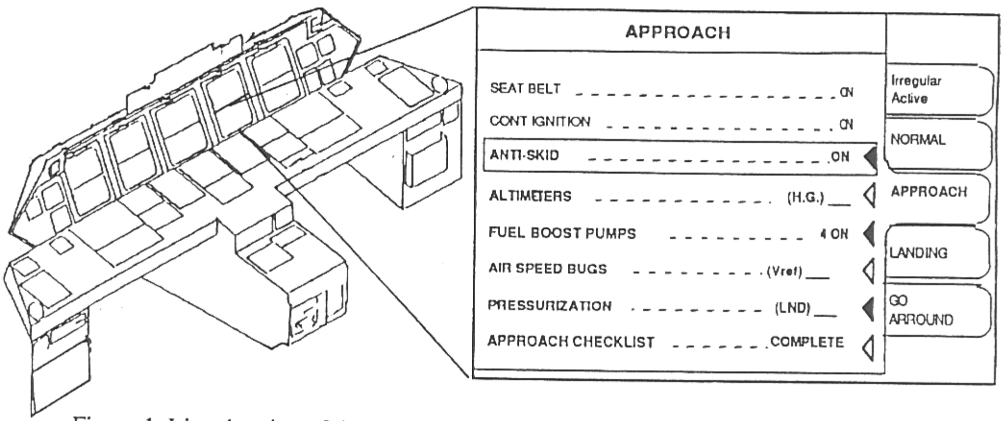


Figure 1. Line drawing of the Advanced Concepts Flight Simulator and a blow up of the manual-sensed checklist. The first two items of the APPROACH checklist are complete.

electronic displays was touch sensitive. When the checklist display was first called up, all checklist items were colored white and the "current-item-box" was situated on the first checklist item. A filled or unfilled triangle, to the right of the checklist item, indicated whether the checklist item was sensed or not (see Figure 1). An unfilled triangle indicated that the item could not be sensed (e.g., *Take Off Briefing* and *Head Count*). A filled triangle indicated an item that could be sensed by the system (e.g., *Flap Position* and *Rudder Trim*). The checklist system automatically displayed the relevant synoptic display for the current checklist item in the upper portion of the display.

Manual-Sensed Checklist. The key difference between the two versions of the electronic checklist was whether the electronic checklist system, or the human, checked the "system-state" first. Figure 1 shows what the "manual-sensed" checklist display would look like part way through the execution of the approach checklist. The first two items are colored green and the triangle symbols are removed to indicate their completion. The third item has been manually acknowledged by the pilot, but the checklist software has sensed it as not accomplished. The item has been marked as skipped (amber) and the "current-item-box" has not advanced to the next item. The checklist will sense the item only after the pilot touches the display to acknowledge completion.

The last item in every checklist was *"Checklist ... Complete."* When a crewmember touched this last item, all skipped or uncompleted checklist items were displayed. The pilot could either return to these items or "override" and continue to the next checklist. The override option was designed to allow the crew to exit a checklist at their discretion even though one or more items were sensed to be incomplete. The checklist design philosophy was to provide reminders, but not to lock-out the crew's control of the situation.

Automatic-Sensed Checklist. In the *automatic-sensed checklist,* all configuration tasks and actions on the systems displays and overhead panel were still performed by the flight crew; however, the actual operation of the checklist was automated as much as possible. When this checklist was called up, all of the sensed items were checked by the system, and items which were sensed as complete were immediately displayed as such. Incomplete and unsensed items were displayed as unaccomplished. If all checklist tasks had been completed and all checklist items were sensed, the only involvement required by the pilot was to manually acknowledge the *"Checklist Complete"* item.

The normal *paper checklists* were printed in standard airline format on a single 8.5 x 11 inch card. The irregular paper checklists were bound in a booklet modeled on Boeing's Quick Reference Handbook (QRH) format.

As part of the experimental design, a procedural manipulation was also introduced, i.e., instruction to perform immediate action items of certain emergency checklists either from memory, without waiting to look them up in the checklist, or by following these items from the checklist. Although the practice of performing certain critical actions immediately, automatically, from memory does allow a quicker response in emergency situations, it also decreases the amount of thought that precedes the action. The time savings gained by this procedure may be overshadowed by an increase in errors if the situation is misinterpreted. This manipulation was expected to affect the handling of the emergency situation described below.

Flight Scenario. This focus of this paper is on the last leg of this simulation, which began in San Francisco (SFO). The crew was told to expect takeoff from Runway 28 Left in San Francisco. On the second leg of the simulation, the crews had received a note (from the "previous crew") advising them that the #1 engine was showing signs of wearing out (although parameters were still within the acceptable range), and recommended that it be watched. Also, just prior to this leg, crews were asked by flight dispatch to monitor the #1 engine, as it had been having some problems, and to log engine parameters during various phases of the flight. The weather advisory included a warning of heavy bird activity sighted off the departure end of runway 28.

Shortly after the aircraft rotated, the flight crew heard a loud "bang." The left (#1) "engine fire handle" lit for 9 seconds and then went off. The EICAS display indicated an "Engine Fire." Engine core speed (N2) and fan speed (N1) indication for the two engines dropped. However, the left engine (#1) recovered (from bird ingestion) to a somewhat reduced thrust. The right engine (#2) did not recover and engine indications showed very low N1, N2, and EPR values. Additionally, a high volume, low frequency sound (i.e., what the crew would hear if the aircraft were actually vibrating) was present. While this was occurring, the crew heard an aircraft on the right parallel runway reporting a bird strike and being routed by Air Traffic Control back to the airport.

At this point, the crew had to determine if the #1 engine was on fire, and, if so, decide whether or not to shut it down. Note that the scenario had "primed" the flight crews for a left engine failure. Additionally, if the electronic checklist crews selected the "ENGINE FIRE" checklist, they saw as the first item, "#1 engine switch off." The combination of events and system indications in this scenario was designed to provide a considerable amount of ambiguity, workload, and distraction, and demanded quick decision making.

Upon completion of each experimental segment, each crew filled out the NASA Task Load Index form (TLX; Hart and Staveland, 1988). This rating form probed the crews on six workload dimensions, and yielded an overall subjective workload rating for each crewmember for each flight segment. Additionally, utilizing videotapes of the simulation, a tally was taken of the available informational items concerning aircraft status that were discussed by the crew. This tally provided an (admittedly imperfect) indication of the amount of diagnostic information crews were considering when making their decision.

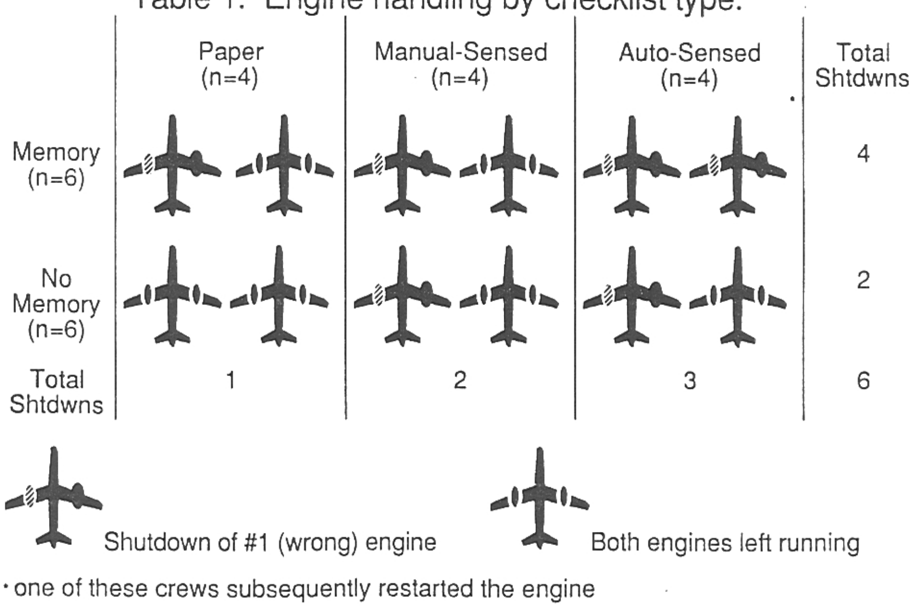
RESULTS

The major dependent variable for this scenario was the handling of the engine problems - specifically, did the crews elect to shut down one of their engines, and, if so, which one? Given the conflicting cues present, it was possible to "justify" shutting down either engine. Although the most salient cues (e.g., the 9-second fire indication and, when present, electronic checklist) had indicated a fire in the #1 engine, engine parameter readings suggested that the #2 engine was actually more severely damaged. Shutting down the #1 engine would have left the crew with only one marginally operative engine with which to return to SFO. A better course of action would have been to leave both engines running (reducing thrust if necessary to reduce "vibration") to ensure enough available power for the climb to terrain clearance altitude, the flight back around to runway heading, and landing. Although the small subject sample in this study makes statistical tests inappropriate, several trends are suggested by the data. These trends may be best illustrated by discussing results in terms of descriptive profiles of crews who took each of the possible courses of action.

Interestingly, *no crews* shut down the #2 engine. This result suggests, at the very least, that crews had been successfully "primed" to expect trouble to be associated with the left (#1) engine. Apparently, the salience of the cues associating the problems with the #1 engine overrode the less salient but more informative indications that the #2 engine was more severely damaged.

As shown in Table 1, half of the crews (6 out of 12) shut down the #1 engine (one crew subsequently restarted the engine). Of these, four were "memory" crews, i.e., crews who had been instructed to

Table 1. Engine handling by checklist type.



· one of these crews subsequently restarted the engine accomplish the immediate action items of the ENGINE FIRE checklist from memory, without looking them up on the checklist. With respect to checklist condition, over half of the crews who utilized a form of the electronic checklist (5 out of 8) shut down the relatively healthy engine, as opposed to one out of the four in the paper condition. Interestingly, neither of the two crews that had to take the time to get out a paper checklist before acting (i.e., paper checklist/no memory condition) made this error.

Crews that erroneously *shut down the #1 engine*, then, tended to be those with electronic checklists, who were initiating the shutdown from memory. Additionally, shutdown crews discussed very little information concerning the status of the aircraft (x=1.2 items). Three of these crews shut down the engine at or below 1300'. Not surprisingly, pilots flying in these crews recorded the highest overall levels of workload of any crewmembers (x=78).

Crews that *left both engines running* tended to be those whose shutdown procedures were less automated - i.e., they were not performing the checklist from memory, and/or were using the traditional paper checklist. These crews also discussed relatively more information concerning the status of the aircraft (x=4.6 items). Mean overall workload ratings for the pilots flying in these crews were the lowest of all crewmembers (x=61). Workload ratings for the pilots not flying did not vary as a function of shutdown decision - mean overall ratings were 70 for shutdown crews and 71 for no shutdown crews.

IMPLICATIONS

Electronic Checklists, Information Transfer, and Decision Making

Crew performance, in terms of operational errors and crew coordination, has been shown to be related to the accuracy of situation assessment, as evidenced by the amount of information transfer among crewmembers (Mosier, 1990). The crew responses in both the "memory" condition and the electronic checklist conditions suggest that information gathering in these conditions was short-circuited by, in one case, the implicit command to accomplish shutdown as quickly as possible, and, in the other case, the compelling (mis–) information that it was the #1 engine that needed to be shut down. In fact, examination of information transfer as a function of checklist type reveals that the mean number of informational items discussed by crews decreased as the checklist became more automated, from x=4.25 (crews with paper checklist) to x=3.0 (manual-sensed) to x=2.25 (auto-sensed).

Salience

The logical conclusion from the results of research on salience effects on decision making has been that, in a diagnostic situation, the brightest flashing light, or the gauge that is largest or most focally located will bias the operator toward processing its diagnostic information content over that of other stimuli (Wickens, 1984). Time pressure, stress, or information overload can cause a "perceptual tunneling" (Stokes, Barnett, & Wickens, 1987) and exacerbate this tendency to focus on central or salient cues.

In one sense, the documentation of the above salience effects supports the need for the checklist, especially in emergency situations. The checklist may serve as a means to focus attention on the important items of information. The introduction of the electronic checklist may enhance this effect, making the *checklist itself* one of the most salient informational items in the cockpit. Therefore, electronic checklists must be designed to be as accurate system sensors as possible, with on-line sensing at a very high sample rate.

An implicit hazard inherent in electronic checklists is that they will become a sort of end in themselves, rather than a means to a good decision. Crew responses in the electronic checklist conditions suggest that both designs encouraged flight crews *to not* conduct their own system checks. Rather, they relied upon the checklist as a primary, rather than back-up, system indicator.

References

Degani, A., & Wiener, E. L. (1990). *The human factors of flight-deck checklists: The normal checklist* (NASA contractor report 177549). Moffett Field, CA: NASA Ames Research Center.

Department of Transport. (1990). *Report on the Accident to Boeing 737-400 G-OBME near Kegworth, Leicestershire on 8 Jan, 1989.* (Aircraft accident report 4-90). London: Air Accident Investigation Branch.

Hart, S. G., & Staveland, L. (1988). Development of the NASA Task Load Index (TLX): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*, pp. 139-183. Amsterdam: North-Holland.

Mosier, K. (1990). *Decision Making on the Air Transport Flight Deck: Process and Product.* Unpublished doctoral dissertation, University of California, Berkeley.

National Transportation Safety Board. (1988). *Northwest Airlines. DC-9-82 N312RC, Detroit Metropolitan Wayne County Airport. Romulus, Michigan. August 16, 1987* (Aircraft accident report, NTSB/AAR-88/05). Washington, DC: Author.

National Transportation Safety Board. (1989). *Delta Air Lines, Boeing 727-232, N473DA. Dallas-Fort Worth International Airport, Texas. August 31, 1988* (Aircraft accident report, NTSB/AAR-89/04). Washington, DC: Author.

National Transportation Safety Board. (1990). *USAir, Inc., Boeing 737-400, N416US. LaGuardia Airport. Flushing, New York. September 20, 1989* (AAR-90/03). Washington, DC: Author.

Palmer, E., & Degani, A. (1991). Electronic checklists: Evaluation of two levels of automation. In R. Jensen (Ed.), *Proceedings of the Sixth Symposium on Aviation Psychology*. (pp. 178-183).Columbus, OH.

Stokes, A. F., Barnett, B. J., & Wickens, C. D. (1987). Modeling stress and bias in pilot decision making. *Proceedings of the Annual Conference of the Human Factors Association of Canada.*

Wickens, C. D. (1984). *Engineering psychology and human performance*. Columbus, OH: Charles Merrill.