

Estimating the Benefits of Human Factors Engineering in NextGen Development: Towards a Formal Definition of Pilot Proficiency

Lance Sherry¹

Center for Air Transportation Systems Research, George Mason University, Fairfax, VA, 22030

Michael Feary²

NASA Ames Research Center, Moffet Field, CA, 94035

Karl Fennell³

United Airlines, Denver, CO

Peter Polson⁴

University of Colorado, Boulder, CO

The concepts-of-operation proposed for the Next Generation Air Transportation System (NextGen) implicitly require a significant improvement over existing standards for flightdeck human-computer interaction. Whereas in today's airspace operations there is no routing penalty for delayed response to a required ATC maneuver, flights in high density NextGen airspace that are unable to respond to off-nominal situations in a timely manner, will lose their slot and be shifted to a downgraded level of airspace resulting in flight delays and/or increased route distance.

Current design and certification processes for avionics, aircraft, and pilots prove the reliability of the "deterministic" automated functions in a comprehensive manner. The design and certification requirements for ensuring and testing the reliability of the inherently "non-deterministic" operator interaction with the automation are not rigorous and are the source of operational inefficiencies and reduced safety margins. Unless the design and certification process are radically modified and refocused, pilots will find themselves with the same types of issues that researchers documented with the introduction of the "glass cockpit" in the 1980s and 1990's.

This paper provides a quantifiable definition of Human Computer Interaction performance and explicit measures of individual and crew proficiency. A method for estimating revenue-service cost savings generated by improved proficiency is described along with an example of the cost savings benefits accrued by a hypothetical large U.S. domestic carrier experiencing improved proficiency in response to FMS error messages (\$45M per year). A discussion of the implications and limitations of the definition of proficiency and the cost savings model is provided.

I. Introduction

CONCEPTS-of-operations proposed for the Next Generation Air Transportation System (NextGen) are designed to increase the effective airspace and airport capacity^{1,2}. Increased capacity will be achieved by more

¹ Director, Center for Air Transportation Systems Research, 4400 University Drive, MS4A6, Fairfax, VA, 22031. AIAA Member.

² Research Psychologist, NASA Ames Research Center, MS: 262-4, Moffet Field, CA, 94035

³ Captain/Instructor, United Airlines Training Department, Denver, CO.

⁴ Professor Emeritus of Psychology and Fellow, Institute of Cognitive Science, Campus Box 344, University of Colorado, Boulder, CO. 80309.

efficient use of the existing infrastructure through trajectory-based operations and reduced separation in high-density airspace. Explicit in these concepts-of-operations are requirements for increased navigation performance. Flights without the capability to meet the navigation performance requirements will be excluded from the airspace.

Implicit in these concepts-of-operations are requirements for maneuver accuracy and responsiveness. The ability to meet these requirements is linked to the *proficiency* of the flight crew in performing the required maneuvers in a timely and robust manner. The current air traffic control paradigm is lenient with regard to flights that do not respond in a timely manner. In the NextGen concepts of operations (e.g. super-density operations), flights that cannot respond to off-nominal situations in a timely manner will be shifted to downgraded airspace and experience flight delays and/or extended distance routing¹.

Meeting this implicit requirement for robust human-computer interaction is not guaranteed. During the 1980's additional functions, colloquially known as "glass cockpit" functions, were added to the airline flightdeck. The tipping-point in pilot confidence in the automation and pilot proficiency was not achieved for two decades after the initial introduction. Researchers extensively documented the issues with the pilot training, certification, and use (or lack of use) of the glass-cockpit in revenue service operations in the in the 1980's and 1990's^{3, 4, 5, 6, 7}. Unless the design and/or certification process is modified, there is little evidence to indicate that the pattern of misuse, disuse, and abuse⁸ will not be repeated.

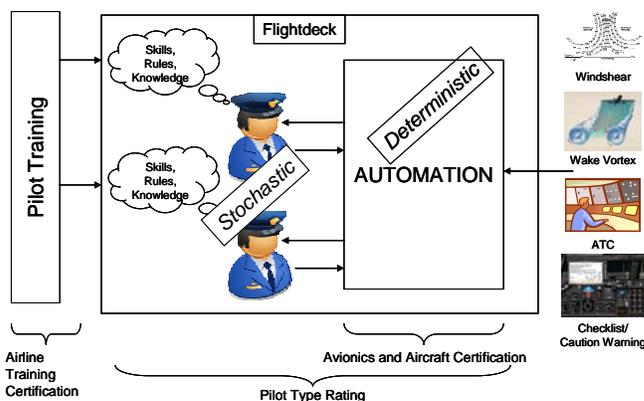


Figure 1. “Deterministic” flightdeck automation is designed, tested and certified to higher standards, than the inherently “stochastic” pilots that are type-rated to respond to emerging situations by executing the appropriate tasks using the correct automation functions.

A fundamental irony is that the current design and certification processes for avionics, aircraft and pilots require specification, analysis and testing of the reliability and performance accuracy of the “deterministic” automation functions. These processes do not apply the same level of rigor to the reliability of the inherently “non-deterministic” human operator. Yet in many mission scenarios, the operator is required to recognize an emerging situation (e.g. windshear, fuel leak, optimum altitudes) and select the correct automation function. See Figure 1.

Following the dictum that “what is measured, gets improved,” this paper provides a framework for a quantifiable definition of flightcrew proficiency for the purposes of design and certification. The paper also provides a model of revenue cost-savings generated from improved flightcrew proficiency. The paper is organized as follows: Section 2 provides a definition of task proficiency that can be used to evaluate

NextGen concepts of operations and a model for estimating the return-on-investment for improved design. Section 3 describes the Cost Model. Section 4 provides a case study application of the cost model. Section 5 includes implications of the cost model.

II. Human-Computer Interaction and Task Proficiency

A. Definition of a Mission Task

Interaction between the human operator and automation on the flightdeck of a commercial airliner occurs when the flightcrew use the automation to complete mission tasks. Mission tasks are the explicit set of tasks required to conduct the mission and include actions such as: executing ATC instructions, performing checklist or procedure items, and responding to caution/warning messages. Airlines have historically provided detailed instructions on how to respond to scenarios in all of these categories in the form of proprietary Standard Operating Procedures (SOPs) that embed the rules and regulations of the Federal Aviation Regulations (FAR) and Aeronautical Information Manual (AIM).

The mission tasks, initiated by voice communication, aural alerting, visual cues, or prospective memory, are performed using flightdeck automation such as the Flight Control Computer and its Mode Control Panel (MCP) or the

<u>Mission Tasks triggered by ATC:</u>	<u>Mission Tasks triggered by Procedures or Checklist</u>	<u>Mission Tasks triggered by FMS Error Messages:</u>
“Climb maintain FL 2-2-0”		
“Proceed direct to waypoint XXX”	Set Altimeter markers to Min Cleanup Altitude with BARO selected	Diagnose a mismatch in fuel sensors triggered by message FUEL DISAGREE – PROG 2/2
“Intercept course 0-1-4 degrees from present position”	Set Engine displays to compact format	Diagnose error in automated procedure tuning triggered by message NAV INVALID – TUNE XXXX –
“For weather, you are cleared to offset your current route 20 nautical miles right. Report when clear of the weather”	Checked & Verbalized Rwy and Final clear	Diagnose flightcrew data entry problems triggered by messages: INVALID ENTRY, NOT IN DATABASE, and INVALID DELETE
“Fly heading 180, join the Gordonsville 060 degree radial, track the radial inbound to Gordonsville”	PM crosschecks altimeters and makes callout: “Crosschecked”	Re-enter data required before takeoff triggered by message: TAKEOFF SPEEDS DELETED –
”Hold west of Boiler on the 270 ⁰ radial. Right turns. 10 mile legs. Expect further clearance at 0830”		

Table 1. Sample of mission tasks. Some mission tasks are generated by ATC, other mission tasks are generated by checklists and error/caution/warning messages.

Flight Management System (FMS) and its Multi-Function Control and Display Unit (MCDU). Table 1 provides a sample of mission tasks performed using the MCP and MCDU.

When a task is triggered, the flightcrew must correctly identify the task and its associated parameters, and then determine which function of the automation to use. For example, to execute the ATC instruction to perform a Hold at a waypoint, the flightcrew should use the FMS “LNAV Hold” function which is explicitly designed to automate the task of flying holding patterns with an accuracy that cannot be achieved flying manually or by using the MCP (especially in the presence of a crosswind). In some cases, mental math is required to convert parameters from the ATC instruction to data that can be entered into the automation (e.g. reciprocal of course).

After deciding which function to use, the flightcrew must *access* the function. In the case of the MCDU this may involve several button-pushes to locate the correct page. Then the flightcrew must *enter* the appropriate data or make the appropriate selections. In most cases, the entry/selection must be “cross-checked” with the other pilot and then “executed” by an additional operator action (e.g. EXECute key on the MCDU). For many tasks, especially those associated with following the flightplan, the airplane trajectory must be monitored closely.

A summary of the operator actions for the task “Hold west of Boiler on the 270⁰ radial. Right turns. 10 mile legs. Expect further clearance at 0830” is shown in Figure 2. For more information on canonical models of human-computer interaction see Sherry et.al^{9,10}.

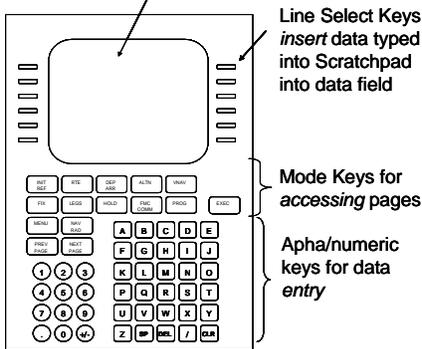
B. Task Proficiency

Task proficiency is the ability to complete a task, such as the “Hold” task shown in Figure 2, within the allowable time window. Different pilots performing the task in various situations will exhibit different times to complete the task.

A typical distribution for time-on-task for a given task is shown in Figure 3. The distribution exhibits a long right-tail. The percentage of subjects that were able to complete the task within the allowable time window are considered proficient. The subjects that were unable to complete the task within the allowable time window are considered to be not proficient. The tail of the distribution in excess of the allowable time window defines the Probability of Failure-to-Complete (PFtC) for the task.

The key parameter is the Allowable Time Window. This determines proficiency. The plans for NextGen will be changing this parameter in a subtle way. First, use of automation and specific procedures will be required for access to certain airspace. More of the airspace will operate like the approach regime does today. The flight will be given access only on condition that a specific type of procedure will be flown. Second, to accommodate the high-density operations, the allowable time window will shrink. The main issue with the shrinking time window is not the nominal procedures. These will be design and tested to ensure pilots can perform them routinely within the safety margins.

MCDU page displays data and accepts *inserted* data



- 1) Id Task as Hold task
- 2) Decide to use the FMS "LNAV Hold: function
- 3) Press HOLD Function/Mode Key
- 4) Press LSK 6L, if a holding pattern is already in the route.
- 5) Line select waypoint identifier for "Boiler" to scratchpad.
- 6) Press LSK 6L.
- 7) Enter the quadrant and the radial into the scratchpad, W/270.
- 8) Press LSK 2L.
- 9) Enter the turn direction into the scratchpad, R.
- 10) Press LSK 3L.
- 11) Enter the leg distance into the scratchpad,10.
- 12) Press LSK 5L.
- 13) Enter expect further clearance time into the scratchpad, 0830.
- 14) Press LSK 3R.
- 15) Verify the resulting holding pattern on the ND
- 16) Press EXECUTE
- 17) Monitor trajectory for entry and following the Holding pattern

Figure 2. Example interaction with MCDU required to execute task for Hold west of Boiler VOR on 270 degree radial.

The issue is the off-nominal or emergency tasks that must be addressed quickly. It should be notes that these tasks occur with low frequency and have performance reliability issues as explained in the next section.

C. Estimating Task Proficiency

Task Proficiency, or it's inverse, the Probability of Failure-to-Complete a Task, for a given task can be determined through usability testing. Operators selected randomly from a population of operators perform a series of tasks under the workload ranges expected in revenue service. This process can be time and cost prohibitive for the schedule and budget of a typical avionics development program.

To overcome the cost and time constraints, alternative analytical methods have been developed. For example, the Human Cognitive Reliability (HCR) model¹¹ provides an estimate of Non-response Probability based on the type of cognitive processing (i.e. reflexive, rule-based, or knowledge-based). Another approach, the Success Likelihood Index Method (SLIM) organizes expert judgments on several categories of factors including: environment, task complexity, training, and individual skills to derive estimates of probability of failure to complete¹². Application of these models for the purpose of NextGen-type design and certification require deep domain knowledge and some expertise in psychology.

Recent developments in the area of "affordable cognitive models" has made feasible the implementation of estimates for expert time-on-task in easy to use tools for software developers and system engineers²². Other recent research in the aviation domain⁹ has duplicated research result on the performance of procedures in the military

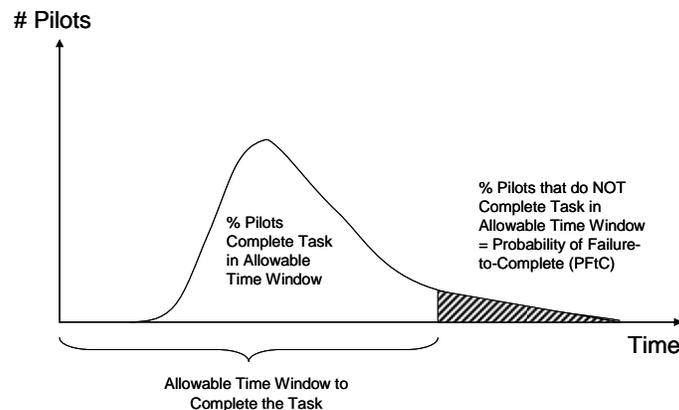


Figure 3. Distribution of Time-to-Complete a Task. The shaded region represents the percentage of pilots that did not meet the Proficiency standard. This defines the Probability of Failure-to-Complete the task.

domain²³ where “the rate of forgetting is predominantly a function of the number of steps ... [and] what tended to be forgotten most were the steps not cues by the equipment or by preceding steps.” A simple, first order model, based on empirical data (Ref 13, 14, 15), captures the relationship between the number of memorized actions required to complete all the decision-making actions and physical actions (e.g. button pushes) in a task, and the Probability of Failure-to-Complete.

$$\text{Probability to Failure to Complete a Task} = 0.1753 * \sum \text{Saliency and Semantic Cue for each Operator Actions}$$

Where the Saliency and Semantic Cue for each Operator Action is:

- 1 if the task is frequent, or if the task is infrequent and the visual cue is within the field of view and exhibits a semantic match to the task and the next action
- 0.25 if the visual cue is within the field of view and exhibits a partial semantic match to the task and the next action
- 0 if the visual cue is not within the field of view, or does not exhibit a semantic match to the task or next action

For purpose of training, estimates of the number of repetitions required to master the execution of a task have been developed¹⁶. Table 2 shows the number of repetitions on each day for subjects learning 30 tasks. These estimates are useful for designing training programs and the design of the training materials (e.g. Computer Based Training) The models are derived from a combination of empirical data (Ref 14, 15) and from simulations of List Learning cognition (Ref 16): These results are consistent with skill acquisition models from other domains such as (Ref 17).

Total Score: Saliency and Semantic Cue for Task	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Total
4	1.5	1.2	1	1					4.7
8	4.5	2.4	1.8	1.5	1.4	1			12.5
12	8.5	3.9	2.7	2.1	1.8	1.6	1		21.7
16	15.5	6.6	4.2	3.2	2.6	2.2	2	1.8	38.2
24	23	9.5	5.9	4.3	3.4	2.9	2	1.8	52.8

Table 2. Estimates of the number of repetitions and days required to achieve mastery of tasks with different levels of complexity. The complexity is defined by the Total Score: Saliency and Semantic Cue for the Task as defined above.

The data in Table 1 identifies the estimated number of repetitions per day required to maintain mastery seven days from the training event. For example, when the Saliency and Semantic Cue for Operator Actions, as determined by the heuristic defined above, is 8, six days of repetitions are required to achieve mastery. Approximately 5 repetitions are required on the first day, followed by 3 on the second, 2 on the third, fourth and fifth days.

D. Individual and Flight Crew Target Level of Proficiency

The Individual Proficiency for a group of operators (e.g. pilots type rated on an aircraft) is shown for a hypothetical set of tasks in Figure 4. The proficiency for each task is determined from the distribution in Figure 2. The tasks, listed from safety critical tasks on the left to efficiency-only tasks on the right, represent the tasks required in response to all plausible mission scenarios.

In a seminal paper titled “How a Cockpit Remembers it’s Speed,” (Ref 18) described how crew members assist each other in completing flightdeck tasks. The improved proficiency exhibited by a two person crew can be modeled to the first-order by the equation shown below. There are several assumptions associated with this model: (i) crew members are performing the task in parallel (or at least are available to assist or cross check each other), (ii) crew are selected randomly from the distribution in Figure 2. (iii) each crew member exhibits a proficiency for a task of p_i .

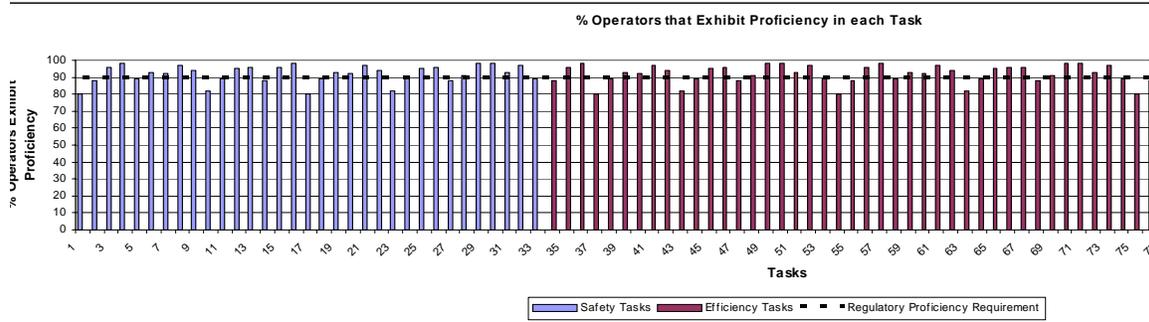


Figure 4. Task Proficiency profile for a population of operators. All tasks required to respond to plausible mission scenarios are ranked by importance to safety (left) and efficiency (right) on the horizontal axis. Bars represent the average number of operators in the population that exhibit proficiency on each task. Dashed-line shows regulatory requirement.

$$\begin{aligned}
 \text{Two Crew Proficiency} &= 1 - (1-p_1) \cdot (1-p_2) \\
 &= 1 - (1 - p_1 - p_2 + p_1 p_2) \\
 &= p_1 + p_2 - p_1 p_2
 \end{aligned}$$

Under the assumption that the crew knowledge and skills are homogeneous, $p_1 = p_2 = p$, the equation for Two Crew Proficiency reduces to:

$$\text{Two Crew Proficiency} = 2p - p^2$$

The relationship between Two Crew Proficiency and Individual Crew Proficiency, illustrated in Figure 5, shows the degree to which “two heads are better than one.” The Two Crew Proficiency increases twice as fast as the Individual Crew Proficiency ($2p$) for $p < 0.7$, then asymptotes to 1 as p increases (and p^2 becomes a dominant term). For example, two crew members drawn randomly from the distribution in Figure 2, with individual proficiency of 70% each, yield a Two Crew Proficiency of 90%. Two crew members with individual proficiency of 90% each, yield a Two Crew Proficiency of 99%.

An important question is the level of proficiency that must be achieved by individual crew members to achieve a Target Level of Proficiency for two crew members. The chart in Figure 4 shows the Probability of Successfully completing all the tasks that are required to respond to all plausible scenarios that can occur in a mission. The tasks on the left of the chart represent the safety critical tasks. The tasks on the right represent the efficiency-only tasks.

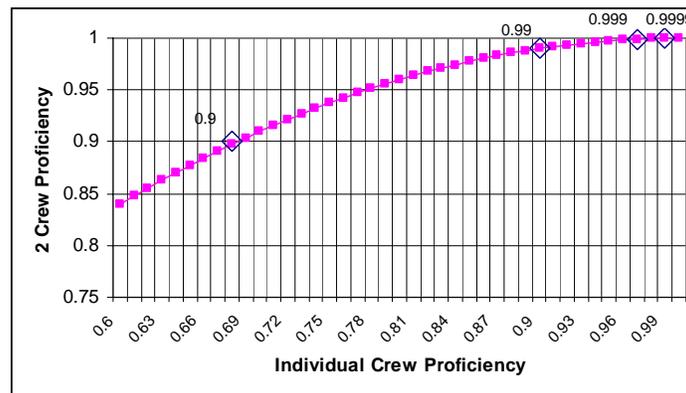


Figure 5. The relationship between Individual Crew Proficiency and the Two Crew Proficiency. Assuming homogeneous proficiency between crew members, to achieve a 90% 2 Crew Proficiency, the individual proficiency of the crew must exceed 70%. To achieve a 99% 2 Crew Proficiency, the individual proficiency of the crew must exceed 90%.

To achieve a Target Level of Proficiency for the safety tasks of 99.9999%, each individual crew member must demonstrate proficiency approaching 99%. To achieve a Target Level of Proficiency for the efficiency-only tasks, the individual crew members need to demonstrate an efficiency of 70%.

E. Families of Tasks and Near-Transfer

Training and testing to proficiency of *all* the safety critical tasks for each pilot would be time and cost prohibitive due to the number of tasks, the number of repetitions required to achieve mastery, and the time to set-up and train or evaluate each task.

To overcome this constraint, the pilot type-rating system takes advantage of the phenomenon of near-transfer. Near-transfer occurs when a skill acquired in practice and/or demonstrated during evaluation close enough to similar skills that credit for the family of skills can be taken¹⁹. For example, demonstration of proficiency of upset recovery at low speeds and at low altitude is sufficient to claim credit for proficiency at upset recovery at other points in the flight regime. Likewise, demonstration of proficiency at performing the Hold task described above is sufficient to claim proficiency in performing all Hold tasks such as Hold at Present Position, Hold at active waypoint, Hold at downpath flightplan waypoint, and Hold at a waypoint not in the flightplan.

In this way the tasks listed in Figure 3 can be grouped into task families. A single task from each family can be selected for training and proficiency evaluation creating a systematic method of ensuring proficiency across all plausible mission scenarios.

The definition of what constitutes a family of tasks and the criteria for grouping tasks is an open area of research.

III. Cost Model for Proficiency

This section describes a cost model for evaluating the revenue-service cost savings yielded by improvement in pilot proficiency. This improvement may be the result of improved user-interfaces, training, or experienced pilots. The model computes the excess costs accrued by a flight as the result of the failure to complete a task. The underlying assumption of this model is that failure to complete a task does not result in an accident, but instead results in a delayed departure/arrival.

Estimates of the cost saving benefits accrued by improved proficiency are a function of the frequency of occurrence of a task, the probability of failure to complete the task (within a time threshold), and the cost of the additional flight time or distance flown.

An estimate of the cost savings benefits can be derived in the three step process as follows:

Step 1: Compute an Estimate of the Annual Excess Costs to Airline Revenue Service Operations:

$$\text{Excess Costs per Flight} = \sum \text{FOC}(i) * \text{PFtC}(i) * \Delta\text{FT}(i) * \text{ADOC}$$

where:

FOC(i) = Likelihood of Task i being performed on a given flight (or Frequency of Occurrence)

PFtC(i) = Probability of Failure to Complete Task i in time t_i

$\Delta\text{FT}(i)$ = Additional Flight Time Incurred by Failure to Complete the Task i (based on Mission Importance)

ADOC = Airline Direct Operating Costs

Daily Excess Costs to Airline = Excess Costs per Flights * Average Number of Airline Flights per Day

Annual Excess Costs to Airline = Daily Excess Costs to Airline * 365 days per year

It should be noted that the failure to complete a task results in additional costs to the airline to reflect the additional effort to avert a safety incident. For example, failure to respond to an FMS Error message indicating a possible fuel leak, would result in a delay in departure or a diversion and not necessarily in an incident or accident.

Step 2: Adjust the PFtC(i) profile to account for improved proficiency

The goal of improved proficiency is to identify and fix tasks that exhibit the potential to have high values (e.g. > 0.3) for the Probability of Failure to Complete (PFtC) the task. Experience in improving pilot proficiency indicates that it is not always possible to eliminate the causes of high values of PFtC for all tasks. There are several reasons that a perfect PFtC cannot be achieved: differences in the “semantic state-space” held by each operator, inadequate physical space on the user-interface to list all the appropriate labels to match the task semantics, onstraints on the space and format of labels and prompts. Preliminary research is underway to better quantify the variances in semantic state-space held by pilots and to develop techniques to group similar functions used by a family of tasks.

For these reasons, for the purpose of cost savings estimation, practical experience suggests that starting with a distribution of PFtC, incremental improvements can be made to move 10% to 25% of the tasks that are in excess of a PFtC threshold of 0.3 to less than 0.3. In general, designers will start with improving the performance for the tasks that exhibit the highest importance to the mission. An alternative approach for the purpose of cost savings analysis (e.g. Section IV below), is to spread the improvement of PFtC equally amongst the classes of importance assigning the improvements to individual tasks on a random basis.

Use the same equations described in Step 1.

Step 3: Compute the Benefits Yielded from Improved Proficiency

Take the difference between the results of Step 2 and Step 1.

$$\text{Estimated Annual Cost Savings} = \text{Annual Excess Costs to Airline (Baseline)} - \text{Annual Excess Costs to Airline (improved proficiency)}$$

The example in the next section demonstrates the application of the cost model.

IV. Case Study: Implications of Proficiency in FMS Error Messages

For the purpose of estimating revenue service cost savings benefits, the distribution of the frequencies and the mission importance for the NextGen messages is assumed to exhibit the same profiles as the previous generation of FMS messages (see Table 2 taken from Ref 10). Further, in the absence of overt improvements in the NextGen development processes, the salience of visual cues on the interface for each message will follow similar patterns as the previous generation of messages. Specifically, 57% of the new messages will not provide flightcrews the underlying causes of the message or guide the subsequent flightcrew actions to respond to message, and that overall, 36% of the new messages will occur very infrequently, exhibit high mission importance, and not be supported by salient visual cues on the interface. For further information on the distributions of the FMS error messages see Ref 10.

Mission Impact (estimate of additional flight time if fail to complete the mission task)	Frequency			
	Very Infrequent (> 100 flights)	In-frequent (= 20 flights)	Occasional (=5 flights)	All the Time
Flight Critical (+30 mins delay)	2	1	-	-
Procedure Critical (10 mins delay)	24	9	5	1
Not Flight or Procedure Critical (< 1 mins delay)	9	1	13	-

Table 3. Distribution of mission tasks generated from 67 error messages for the B777 FMS. See[23]

In addition to the assumptions above, the Total Flight Delays, Total Excess Costs per Day, and Total Excess Costs per Annum for a hypothetical large domestic carrier are calculated with the following inputs:

- Average Flights per Day – 3745 (535 aircraft that average 7 flights per day)

- Average hours utilization per aircraft – 13 hours with average flight distance – 630 nautical miles
- Average Direct Operating Cost - \$35 per minute. Biased by ratio of gate-taxi airborne operations (Bureau of Transportation Statistics²⁰).

Table 3 shows the estimate of the flight delays and their associated excess costs accrued by the airline due to persistent interaction with the error messages. The baseline for the study, shown in the left column, is for error messages that exhibit distributions of frequency, importance, and visual cueing similar to the existing FMS message set. The subsequent columns show the excess costs accrued by the airline when 10% and 25% of the error messages with PFtC greater than 0.3 are improved to a PFtC of less than 0.3.

	Baseline (Traditional Design Methods)	Improved Proficiency yields 10% less Tasks with PFtC > 0.3	Improved Proficiency yields 25% less Tasks with PFtC > 0.3
Daily Total Flight Delays	583 hrs	525 hrs (-58 hrs)	437 hrs (-146 hrs)
Daily Total Excess Cost	\$1.2M	\$1.1M (\$-0.122M)	\$0.92M (\$-0.306M)
Annual Total Excess Costs	\$447M	\$402M (\$-44M)	\$335M (\$-112M)

Table 4. Summary of Total Excess Cost per Day and Total Excess Cost per Year resulting from additional flight times accrued by lack of proficiency in the response to new messages to support NextGen procedures. The new messages exhibit historic distributions for frequency and importance to mission, but through improved proficiency, exhibit reduced Probability of Failure-to- Complete

For the airline route properties described above, the hypothetical airline could accrue between \$122K and \$306K in daily savings from the 10% and 25% reductions in Probability of Failure-to-Complete tasks triggered by these error messages. Annual savings range from \$44M to \$112M for the 10% and 25% improvement conditions.

V. Conclusions and Future Work

This paper provides a formal definition of pilot and flightcrew proficiency and provides the framework to evaluate proficiency. Proficiency is defined as the ability to complete a task in the required time window. Proficiency across tasks is split between proficiency for safety related tasks and proficiency for efficiency-only tasks.

The relationship between individual proficiency and flightcrew proficiency is developed. In a two-person flightcrew, individual proficiency levels of 70% are satisfactory for a flightcrew proficiency of 90%. Individual proficiency levels of 90% are required to achieve a flightcrew proficiency of 99%. The concept of families of tasks is introduced as a means to train and evaluate proficiency in a sub-set of tasks that represents the full population tasks by taking advantage of the phenomenon of near-transfer training..

The paper also describes a methodology to estimate savings in Flight Delays and in Airline Direct Operating Costs (ADOC) achieved by efficient use of the functions of flightdeck automation. In particular the savings are generated: (1) by reduced flight-time (including fuel-burn and crew costs) that occur when flightcrews are able to respond to ATC instructions, checklist items, and caution/warning/error messages in a timely manner, and (2) when flightcrews are able to respond to ATC requests to expedite operations or control accurately to a crossing restrictions using advanced functions in flightdeck automation. Both of these circumstances result in reduced distance flown and greater flexibility in on-time performance (i.e. slowing down is better than having to make up time). The underlying assumption of this approach is that NextGen airspace will be operated on the basis of equipage and the ability to perform required tasks. Flights that are unable to perform the required maneuvers will be shunted to alternate airspace and will be required to absorb additional costs.

The methodology described computes the expected value for cost savings based on the: (i) frequency of occurrence of the task, (ii) average cost of failing to complete the task (in minutes flight delay), and (iii) the Probability of Failure-to-Complete the task. The case study illustrates the application of the cost model to airline revenue-service operations and shows that annual savings for a large domestic carrier can be estimated conservatively between \$40M and \$110M.

Acknowledgments

This project was funded by a grant from NASA – Aeronautics – Intelligent Integrated Flightdeck program, and by internal George Mason University Foundation Funds. Thank you for technical assistance and suggestions from: Dan Boorman, Randall Mumaw (Boeing), John Otiker (SWA), Marilyn Blackmon (University of Colorado), Andrea Knight, Ricardo Prada (Google), Karlin Toner (FAA), Robert Mauro (University of Oregon), Immanuel Barshi (NASA-ARC), Mike Valencia, Midori Tanino (FAA), Mike Matessa (Alion Inc.), Michael Gordon (CCPace), Ken Goldwasser (SEER Tech), Tarik Ward (NASA-JSC), Chip Hathaway, Terry Thompson (Metron Aviation), Debra Shreckenghost (TRACLabs), John Shortle, Rajesh Ganesan, John Ferguson, Guillermo Calderon-Mesa, Jianfeng Wang, Loni Nath (CATSR/GMU). Thank you for support of the research to Steve Young (NASA), Amy Pritchett (NASA).

References

- ¹ JPDO (2008) NextGen Integrated Work Plan Version 1.0. <http://www.jpdo.gov/iwp.asp>
- ² SESAR Consortium (2008) SESAR Master Plan D5. MGT-0803-001-01-00. <https://www.atmmasterplan.eu/http://prisme-oas.atmmasterplan.eu/atmmasterplan/faces/index.jspx>.
- ³ Wiener, E., D. Nagel (1989) *Human Factors in Aviation*. Academic Press : NY, NY.
- ⁴ Federal Aviation Administration. (1996) Human factors study team report on the interface between flightcrews and modern flightdeck systems. Washington, D.C. 1996.
- ⁵ Billings, C.E. *Aviation automation: The search for a human centered approach*. 1996. Lawrence Erlbaum Associates: Mahwah, N.J, Inc
- ⁶ BASI: Advanced Technology Safety Survey Report. *Flight Safety Digest Special Issue*. Flight Safety Foundation, June-Aug 1999, pp. 137-216. Available at www.basi.gov.au/sprog/advtek.htm (9/00)
- ⁷ Feary, M. M. Alkin, P. Polson, D. McCrobie, L. Sherry, E. Palmer (1999) Aiding Vertical Guidance Understanding. Air & Space Europe Volume 1, Issue 1, January-February 1999, Pages 38-41.
- ⁸ Parasuraman, R., V. Riley (1997) Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, Vol. 39, 1997
- ⁹ Sherry, L., M. Medina, M. Feary, J. Otiker (2008) Automated Tool for Task Analysis of NextGen Automation. *Proceedings of the Eighth Integrated Communications, Navigation and Surveillance (ICNS) Conference (2008)*, Bethesda, MD. May 5-7, 2008
- ¹⁰ Sherry, L., K. Fennell, M. Feary, P. Polson (2006) Human-Computer Interaction Analysis of Flight Management System Messages. *Journal of Aircraft*. Vol. 43, No. 5, September-October 2006
- ¹¹ Hannaman, G.W., A.J. Spurgin, & Y.D. Lukic (1985) A Model for Assessing Cognitive Reliability in PRA Studies. In *Proceedings IEEE Conference on Human Factors and Power Plants*. Monterey, California.
- ¹² Embrey, D.E. (1984) SLIM-MAUD: Approach to Assessing Human Error Probabilities Using Structured Error Judgement, NUREG/CR-3518, Vol 1. U.S. Nuclear Regulatory Commission, Washington, D.C.
- ¹³ Fennell, K., L. Sherry, R. Roberts. Accessing FMS functionality: The impact of design on learning. NAS/CR-2004-212837. 2004
- ¹⁴ Boorman, D.J., R. J Mumaw (2007) A New Autoflight/FMS Interface: Guiding Design Principles Air Canada Pilots Association, Safety Conference, 2007 acpa.ca
- ¹⁵ Prada, L. Ricardo; Mumaw, R J.; Boehm-Davis, D. A.; Boorman, D.J. (2007) Testing Boeing's Flight Deck of the Future: A Comparison Between Current and Prototype Autoflight Panels Human Factors and Ergonomics Society Annual Meeting Proceedings, Aerospace Systems , pp. 55-58(4)
- ¹⁶ Matessa, M. & P. Polson (2005) List Model of Procedure Learning. *NASA/TM 2005-213465*
- ¹⁷ Bovair, S., Kieras, D.E., & Polson, P.G. (1990). The acquisition and performance of text-editing skill: A cognitive complexity analysis. *Human Computer Interaction*, 5, 1-48.
- ¹⁸ Hutchins, Edwin(1995)'How a cockpit remembers its speeds',*Cognitive Science: A Multidisciplinary Journal*,19:3,265-288
- ¹⁹ Bhavnani, S.K., Peck, F.A., and Reif, F. (2008) Strategy-Based Instruction: Lessons Learned in Teaching the Effective and Efficient Use of Computer Applications. *ACM Transactions on Computer-Human Interaction*, 15,1 (2008), 2:1-2:43.
- ²⁰ Bureau of Transportation Statistics (2008) <http://www.bts.gov>
- ²¹ Sherry, L. & M. Feary. Task design and verification testing for certification of avionics equipment. In *Proceedings of AIAA/Digital Avionics Systems Conference, 2004*. Salt Lake City, Utah. October, 2004.
- ²² John, B. E. & Salvucci, D. D. (2005) Multi-Purpose Prototypes for Assessing User Interfaces in Pervasive Computing Systems. *IEEE Pervasive Computing* 4 (4), 27-34.
- ²³ Druckman, D., & Bjork, R. A. (Eds.). (1991). *In the mind's eye: Enhancing human performance*. Washington, DC: National Academy Press.