Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
April 30, 2015
A Regional Conference of the Society for Industrial and Systems Engineering

Design of a Procedure Analysis Tool (PAT) for the FAA Human Factors Certification Process

Heather Johnson, Houda Kerkoub, Shaida Latifi, Mumtahina Mahmud, and Alexis Nittoli
George Mason University

Corresponding authors' Emails: hjohns17@masonlive.gmu.edu, hkourdal@masonlive.gmu.edu, slatifi2@gmu.edu, mmahmud@gmu.edu, anittoli@masonlive.gmu.edu.

Author Note: Senior Design students at George Mason University.

Abstract: Equipment on airliners are certified by FAA to comply with safety standards. This process is currently accomplished by both inspection and Human-In-The-Loop (HitL) testing – a time-consuming and costly process. A new Regulation FAR25.1302 requires analysis of all procedures; increasing certification costs beyond current capacity.
The Procedure Analysis Tool (PAT) described below is a decision support tool designed to meet FAR25.1302 requirement. PAT simulates pilot performance time on the device under evaluation. The output of PAT is a frequency distribution showing the number of pilots versus time to complete one procedure. The percentage of pilots performing the procedure in excess of an allowed threshold represents the Probability of Failure to Complete (PFtoC) the procedure. Procedures with long right tails are flagged for HitL testing.
PAT was demonstrated on 14 procedures to assess the Multifunction Control Display Unit (MCDU). Analysis shows that the PAT reduces the evaluation cost by 70% per function.

Keywords: Procedure Analysis Tool, Cockpit Design, Human Error Identification, Predictive Modeling.

1. Introduction

The Federal Aviation Administration (FAA) is the United States government agency with responsibility for airworthiness certification of all classes of aircraft and the devices (i.e. equipment) used on the aircraft. To achieve this goal, the FAA coordinates the legislation of Federal Aviation Regulations (FARs) that establish the performance standards that devices are required to meet before they can be approved as airworthy. The FAA works closely with device manufacturers (known as "applicants") to perform the tests and evaluations that show that the devices meets the performance standards in the FARs. To manage the workload and ensure the right expertise, the FAA designates individuals as Designated Engineering Representatives (DERs) to perform the evaluations. In some cases, the DER may be employed by the applicant.

As the outstanding safety system in place for aviation has one-by-one eliminated accident causes, one of the most significant factors contributing to accidents that has risen to the top is "flight-crew error". Human errors account for 56% of the yearly causes of air accidents (Sherry, Feary, Fennell, & Polson, 2009). This is the case where a pilot makes an incorrect action, or more likely, the pilot fails to take a required action.

Unlike the other characteristics of device performance, such as temperature range, vibration, mean-time-between failures, etc. that are specified by the performance standards in the FAR and evaluated for airworthiness, the human factors regulations have historically been made open-ended. First, the human factors regulations are found buried in FARs for specific types of devices (e.g. displays, flight controls). Second, the rules focus on low level human factors issues such as the salience of displays in sunlight, the tactile feel of adjacent knobs, and the ergonomics of input device placement relative to pilot seating position. These rules do not specify performance standards related to procedure performance.

To address this shortcoming, the FAA and the European Aviation Safety Agency (EASA) jointly released a new regulation FAR/EASA 25.1302 – relating to airworthiness standards on installed systems and equipment for use by the flightcrew. This regulation collects all human factors regulations in one FAR. It also significantly strengthens the rules by requiring that in addition to the lower-level (ergonomic) human factors requirements that "the applicant must show that […] flightcrew […] can safely perform **all** of the Procedures associated with the systems' and equipment's intended functions" (FAA, 2013).

The complementary Advisory Circular (AC) 25.1302 provides guidelines for the design and Methods of Compliance (MOC) for the installed equipment on airplanes to minimize errors by the crew and enables the crew to detect and manage errors that may occur (FAA, 2013). This AC provides recommendations for the design and evaluation of controls, displays, system behavior, and system integration that are all part of human factors considerations. Specifically, requirement b-2 calls for the implementation of accessible and usable control and information by the flight crew in a manner consistent with urgency, frequency, and task duration (FAA, 2013).

One of the means of compliance prescribed by the AC is referred to as "Evaluation". Evaluation may be conducted via Inspections or by Human-in-the-Loop testing on one of the following platforms (1) mock-up, (2) on a bench, (3) in a laboratory, (4) simulator, (5) or aircraft (FAA, 2013)

Given the approved means-of-compliance (i.e. Human-in-the-loop testing) with the requirement to evaluate all Procedures, the human factors evaluation of FAR 25.1302 becomes cost and time prohibitive. For example, for a device with 600 functions, the pre- FAR 25.1302 evaluation of a sample of procedures on 30 functions (only) would suffice and require 85 hours of HitL testing and cost approximately $14,500. With the new FAR 25.1302, the HitL testing would take an estimated 1700 hours and cost approximately $989,000 (i.e. 68 times more expensive).

There are other issues with the current human factors process. First, the AC25.1302 describes Evaluation as "assessments of the design conducted by the applicant (or their agent), who then provides a report of the results to the FAA" (FAA, 2013). The three stakeholders mentioned in this sentence are the primary stakeholders. They are (1) the FAA represented by the FAA inspector, (2) the applicant – that is the device manufacturer, and (3) the agent, or the Designated Engineering Representatives (DER) employed by the manufacturer. In the certification process, the DER reports to both the FAA inspector and to the aviation manufacturer. On the other hand, since FAA and the manufacturer have distinct objectives in the certification process, it may create a conflict of interest, tensions, and pressure on the DER. Second, the devices are isolated while certified, yet they are integrated in the flightdeck during flight operation. A more accurate human factors evaluation would assess the performance of the device in the way it is used i.e.: in conjunction with the other flightdeck systems. Lastly, some elements of the current certification process rely on subjective assessments (e.g. workload) and inter-rater reliability can become an issue.

2. Problem Statement

The new regulation FAR25.1302 requirement to evaluate all Procedures is cost and time prohibitive if it has to be accomplished by Human-in-the-Loop testing. The costs and time required to evaluate will grow to a significant percentage of the overall cost development and certification of the device. Also, there is the difficulty faced by the DER, the issue with the integration in the flight deck procedures, and the subjective nature of evaluation.

3. Proposed Solution

The proposed solution is to provide a simulation of flightcrew human performance performing procedures on a device. The simulation is a system engineering model of Time-to-Complete procedures (TTC) based on time-to-complete distributions for each operator action in the Procedure (e.g. button push...) (Patton, 2010). The automation of the evaluation is significantly faster and cheaper, it eliminates the conflict of interest between the stakeholders, it performs integrated flightdeck analysis, and it provides an objective means of analysis. A high level idea of the PAT is illustrated in the Figure1 (Sherry, Feary, Fennell, & Polson, 2009) below where the DER evaluates a proposed system design (device) using a combination of procedure decomposition and operator performance models to generate time-on-procedure distributions for each procedure (Tollinger, 2005).

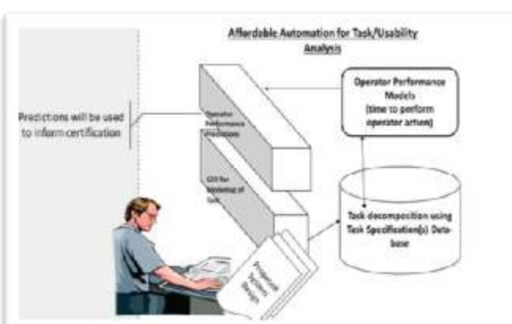

Figure 1. Concept of Operations

3. Approach

The airline Standard Operating Procedures (SOPs) and Computer Based Training (CBT), define the flightcrew procedures and tasks that are performed on the device (Patton 2010) A device displaying good human factors is expected to enable the operators to complete their tasks in the allotted time. Failure to complete the procedure occurs when the procedure cannot be competed in the operational window allowed (Sherry, Feary, Fennell, & Polson, 2009). A distribution for Time to Complete the Procedure (TTC) for a sample of flightcrew in shown in Figure 2. The Probability of Failure to Complete (PFtoC) the procedure in the right tail of the distribution is an indication of whether the device passes the certification test. A procedure showing high PFtoC will require additional HitL testing.

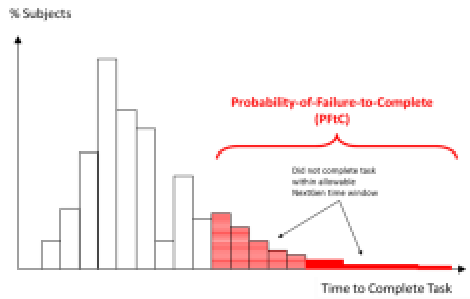

Figure 2. Example Time to Complete the Procedure

4. Description of Tool

The PAT models the timing of pilot operations in the cockpit environment. The input to the PAT is a detailed description of the procedures from the SOP and/or the CBT. The Procedures are listed as a sequence of Operator Actions. The PAT allows the DER to assign a category of Operator Action to each Operator Action. Time distributions for each category are stored in the Human Performance database.

The tool works in six steps as follows:

1. Decompose procedure into OA
2. Categorize into OAC and associate with Human Factors metrics
3. Simulate procedure n times (Monte Carlo simulation with 500 replications)
4. Request threshold for PFtoCT from user
5. Calculate PFtoCT to determine devices requiring HitL testing.
6. Flag procedures requiring further HitL testing

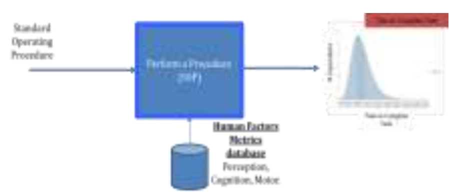

Figure 3. Conceptual Model

The first step is performed using a sequence diagram (see Figure 3). Each OA falls under an Operator Action Category (OAC) associated with a statistical distribution as shown in Table 1 below. After the identification of the OAC for each OA, random numbers are generated according to their statistical distributions before being aggregated to become the procedure time. This operation was performed 500 times using a Monte Carlo simulation to result in a distribution as shown above in Figure 2.

4.1 Mathematical models/algorithms

Two attempts using closed form equation were considered to determine the PFtoCT (1) is the determination of the parameters of the sum of multiple distributions and (2) is the use of convolution. The first option was rejected because it required all the random variables to be of the same type, and the second option was also rejected because it required the time *t* at which the PFtoCT is calculated to be set beforehand. It was then determined that simulation is necessary to compute the parameters of the final distribution including the PFtoC. To give an idea of the algorithm used for this operation, a high level description is shown below:

Comment [H1]: Does that answer the "explain" comment?
Check with Dr. S

Comment [H2]: explain

```
carry out any initializations required.
while not reached end of file do
 {
  read in the OAC for the OA in sequence.
  generate next random number as per OAC.
  add the number to the accumulated sum.
  }
identify the type of distribution.
evaluate the average.
evaluate the standard deviation.
get user input for Threshold of probability of failure to complete.
evaluate the probability of failure to complete the task.
```

In the above algorithm, "random number" refers to the random number generated per the corresponding OAC's statistical distribution. The OA are modeled in series and their sum is illustrated by Figure 4 below demonstrating a simplified version of a Procedure composed of 3 OAs.

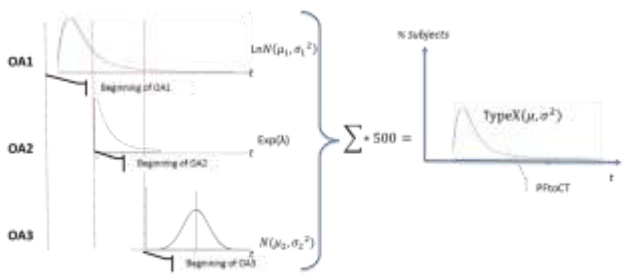

Figure 4. Sequential OA Distributions

4.2.1 Human Performance Data-base of OAC:

There are 17 OACs in total. Each OAC is linked to a statistical distribution. The statistical distributions, show in Table 1 below, and their parameters were either taken out of literature review (ex: Visual Cue is N~ (0.632, 0.011) (Cao, 2013) , or produced by designing experiments using GMU students as subjects.

Operator Action Category (OAC)	Distribution
button push	Normal(0.3, 0.01)
decision (TSL Id)	Normal(0.5,0.002)
decision/choice	Lognormal(2,1)
flight controls manipulation (feet)	Normal(2.31, 0.651)
flight controls manipulation (hands)	Lognormal(1.12, 0.231)
listen to audio (ATC Clearance, Aural Warning)	exponential (1, 0.1)
long term memory item retrieval	Lognormal(0.1, 0.001)
MCC callout	Normal(2,1)
MCC readback	Normal(1.5, 0.2)
Talk to ATC (including Readback)	exponential (0.9, 0.1)
thrust levers manipulation	Triangular(0.1,2,3.5)
TSL Select	Normal(0.1,0.002)
visual check	Normal (1, 0.5)
visual cue: Routine	Normal(0.632, 0.011)
visual cue: Moderate frequency	Normal(1, 0.02)
visual cue: Rare frequency	Normal(1.632, 0.03)
working memory item retrieval	Lognormal(0.1, 0.05)

5. Case-Study of 14 SOP Procedures

For the preliminary analysis, a pool of 14 procedures[1] was defined for the Multi-Function Display Unit (MCDU). In an effort to make this first sample a representative one, procedures were drawn across all phases of flight. For example, the "Initialize Position" procedure would be performed during the "Pushback/Taxi/Take-off" phase. The distribution of procedures per phase of flight is as follows: 4 Procedures for Flight Planning, 3 for each of the Pushback/Taxi/Take-off, Domestic/Oceanic Cruise, and Descent/Final Approach phases, and 1 Procedure was analyzed for the Taxi and Arrival phase.

Each Procedure was (1) decomposed, (2) analyzed for OAC weight, (3) analyzed for statistical distribution fit, and (4) analyzed for PFtoC the Procedure.

An example task description for the Initialize Position task is shown in Figure 5. This procedure had 55 operator actions. The OACs were as follows: (47.27%) "Visual cue", (16.36%) "Long-term memory item retrieval", (12.73%) "Decision/choice", (10.91%) "Button push" and (5.45%) "Working-memory item retrieval" while (1.82%) of the procedure is either "Procedure Identification", "Decision/Choice", "TSL Select", or "Visual Check".

The distribution best fitted a Lognormal distribution with parameters 50 + LOGN (166, 238), and (4) the PFtoCT was calculate to be 1.6%. With a maximum allowable PFtoCT of 2.5% the procedure was estimated to have passed the certification evaluation.

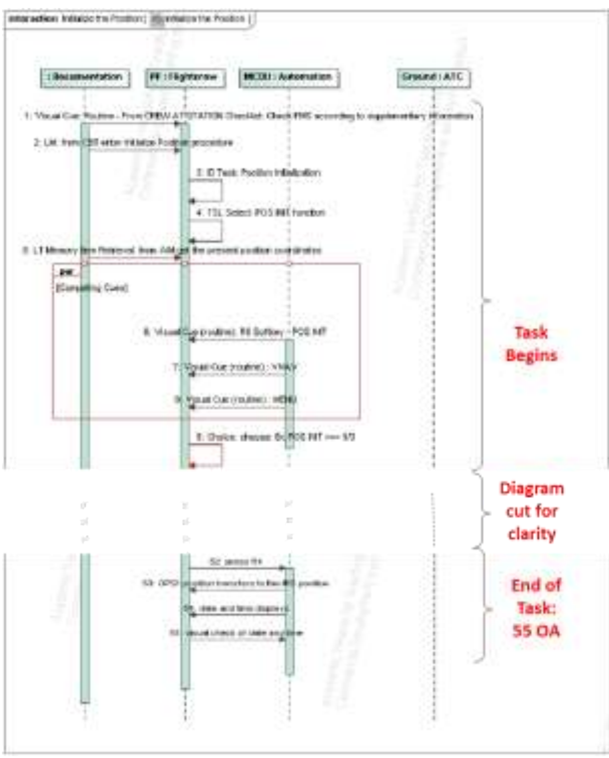

Figure 5. Initialize Position Example Decomposition

The PAT ran 500 replications for each of the 14 Procedures analyzed. The table below summarizes the PFtoC, along with data entry and run times. For demonstration purposes (and lack of user input), the threshold for PFtoC was set to 3

[1] The final results are planned to include at least 30 Procedures to enable comparison between pre-FAR and proposed solution time and budget.

standard deviations from the mean, and the limit for acceptable PFtoCT was set to 2.5%. In this case, three procedures resulted in a PFtoCT above the threshold.

Table 2 Summary of Results

#	Task	CBT Task Time [sec]	Entry time in MagicDraw [h	Run Time Excel [sec]	mean	std	3std from mean	PFtoC (linked)
1	Initialize the Position	27	1:05	3.76	238.0359	192.1603	814.5167	1.60%
2	Enter Company Flightplan	23	1:30	3.21	218.8321	174.5325	742.4296	2.60%
4	Check Flgihtplan Progress	42	1:00	3.66	216.9292	327.774	1200.251	0.00%
5	Switch MCDU ON	47	0:15	19	91.55241	98.1663	386.0513	3.40%
6	Enter Runway and SID info	33	0:30	16	144.5271	144.5271	578.1085	1.00%
7	Enter flightplan manually	51	1:28	13	207.8871	175.1326	733.2848	2.60%
8	Hold at Present Postion	17	2:30	10	238.0359	192.1603	814.5167	1.50%
9	Change turn direction, speed, and check inbound course	25	0:21	14	135.1307	132.722	533.2966	1.60%
10	Check Climb/Cruise Settings before departure	17	0:23	13	107.4081	103.7563	418.677	1.80%
11	Change Speed and Altitude of a non Active waypoint on the Climb	36	0:36	13	89.34417	61.8645	274.9377	2.20%
12	Select Arrival and Approach Procedures	28	0:47	14	170.141	122.9245	538.9146	2.20%
13	Enter a Runway Extension	20	0:36	13	89.58524	98.62805	385.4694	2.40%
14	Enter Hold ExitTime	44	0:26	13	214.7787	164.5245	708.3522	2.40%
15	Select Speed Restriction	17	0:40	15	104.6379	144.9548	539.5024	0.4%

6. Business Case Analysis FOR PAT

Performing the certification as required per FAR25.1302 was estimated to cost $1630/function. Using PAT for certification costs $490/function. That is a 70% saving per evaluated function. The PAT is proposed to sell under three business models: (1) a tool purchase model priced at $300k per unit in addition to a yearly recurring fee, (2) a license purchase model priced at $75k per license in addition to a yearly fee, and (3) an “A La Carte” model priced at $160 per function with a limit of 200 functions per year. The market analysis identified 10 potential buyers. The Return on Investment was calculated to be ROI=190% after 2.5 years, and the breakeven point to be after 1.1 years.

7. Conclusions and Recommendations

The PAT was demonstrated to work on 14 Procedures on the MCDU device. The run time averaged around 12 seconds, and analysis showed that PFtoC ranged between 0.4% and 2.6%. Also, 3 over 14 Procedures (21%) were flagged for HitL testing, and the Visual Cue category gathered the greatest OAC percentage. This preliminary analysis highlights the importance of semantic cues to recognize emerging mission situations which has the greatest effect on the final time distribution.

9. References

Cao, S. (2013). Queueing Network Modeling of Human Performance in Complex Cognitive Multi-task Scenarios.

FAA. (2013). *Advisory Circular 25.1302.*

Patton, E., & Grey, W. (2010, August). *SANLab-CM: a tool for incorporating stochastic operations into activity network modeling.* Retrieved January 31, 2015, from http://www.ncbi.nlm.nih.gov/pubmed/20805610

Tollinger, I., Lewis, R., McCurdy, M., Tollinger, P., Vera, A., Howes, A., Pelton, L. (2005). *Supporting efficient development of cognitive models at multiple skill levels: exploring recent advances in constraint-based modeling.* Retrieved February 21, 2015

Sherry, L., Bonnie, J., & Polson, P. (2010). *System Desing and Analysis: Tools for Automation Interaction DEsign and Evaluation Methods.*

Sherry, L., Feary, M., Fennell, K., & Polson, P. (2009). Estimating the Benefits of Human Factors Engineering in NextGen Development: Towards a Formal Definition of Pilot Proficiency.