Design of a Procedure Analysis Tool (PAT) for Affordable Aviation Device Human Factors Certification

Houda Kerkoub, Heather Johnson, Shaida Latifi, Mumtahina Mahmud, and Alexis Nittoli
George Mason University hkourdal, hjohns17, slatifi2, mmahmud, anittoli@masonlive.gmu.edu

Abstract - **Equipment on airliners must be certified by the FAA to ensure compliance with safety standards that are designed to minimize design-related flight crew errors. This process is currently accomplished by both Inspection and Human-In-The-Loop (HitL) testing – a time-consuming and costly process. A recent Federal Aviation Regulation (FAR) 25.1302 newly requires the analysis of all tasks increasing certification costs beyond current staffing, budget and timelines. The Procedure Analysis Tool (PAT) described in this paper is a decision support tool designed for use by inspectors performing the certification test to meet the FAR25.1302 requirements. The PAT simulates the performance of human operators performing a task on the device under evaluation and may be used as a screening tool to identify tasks that warrant HitL testing, allowing for evaluation within reasonable time and budget.**

The PAT calculates the percentage of pilots that perform the procedure in excess of an allowed threshold representing the Probability of Failure to Complete PFtoC the procedure. Procedures with long right tails are flagged for full HitL testing. PAT was demonstrated on 15 Swiss European airlines Standard Operating Procedures (SOP) to evaluate the Multifunction Control Display Unit (MCDU). Three procedures resulted in a PFtoC above the threshold, and were therefore flagged for HitL testing. This preliminary analysis highlights the importance of semantic cues when recognizing the emerging mission situations which has the greatest effect on the final time distribution. Entering the 15 procedures into the tool took 5.1 hours and 11.56 seconds average runtime. Analysis of the human factors certification process shows that the PAT reduces the evaluation time by 70% per function.

Index Terms - Federal Aviation Regulation, Human in the Loop, Procedure Analysis Tool.

INTRODUCTION

In 2003 Boeing published statistics indicating that there were 139 commercial jet aircraft accidents world-wide between 1993 and 2002 [1]. In 67% of these, flight crew error was cited as a major factor by the investigating authority [1]. These statistics remain consistent across accident categories that have been recently examined [1].

Among the factors identified to have contributed to these accidents is design [1]. Such an analysis led FAA to take measures to address the design and certification of transport category flight decks [1]. As a result, and in a harmonization effort with the European authorities, FAA issued in 2013 a Federal Aviation Regulation (FAR) 25.1302 relating to airworthiness standards on installed systems and equipment for use by the flightcrew [2]. The proposed requirements in FAR25.1302 augments regulations with more explicit requirements for design attributes related to managing and avoiding flight crew error [3]. This regulation collects all human factors regulations in one FAR and adds to the lower-level (ergonomic) human factors that "the applicant must show that […] flightcrew […] can safely perform all of the procedures associated with the systems' and equipment's intended functions" [4] (noting that 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). Although this regulation improves the process, it increases certification cost and time budgets beyond current capacity.

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 [5]. First, the human factors regulations are found buried in FARs for specific types of devices (e.g. displays, flight controls) [5]. 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 [5]. These rules do not specify performance standards related to procedure performance [5].

The complementary Advisory Circular (AC) 25.1302 provides guidelines for the Methods of Compliance (MOC) with FAR25.1302 [4]. 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 [4] .

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 [4].

Given the approved means-of-compliance (i.e. Human-in-the-Loop testing) with the requirement to evaluate all tasks, 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. about 68 times more expensive for a device with 600 functions). Additionally, some elements of the current certification process rely on subjective assessments (e.g. workload) where inter-rater reliability can become an issue.

Since FAR25.1302 focuses on task performance, the tasks are drawn from the Standard Operating Procedures (SOP) (a set of instructions documenting how to perform routine and emergency activities to ensure safety, consistency, and quality). SOPs are found in company manuals as well as training manuals including Computer Based Training (CBT). For this project, Swiss European airlines SOPs on the Multi-Function Display Unit (MCDU) were used to test case the proposed tool.

The Procedure Analysis Tool (PAT) described below is a decision support tool for use by the Designated Engineering Representative (DER) performing the certification evaluation. It is designed to meet FAR25.1302 requirements by enabling analysis of all tasks in a manner consistent with urgency, frequency, and task duration. Since the aircraft cockpit is a dynamic environment, PAT performs the human factors assessment based on one metric: time – a device with good human factors engineering simply enables safe performance of tasks within time. The 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 (TTC) one procedure [6]. 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 [7], as shown on Firgure1 below, are flagged for HitL testing.

The PAT enables decomposition of a procedure into finite Operator Actions (OA) using the Task Specification Language (TSL) framework [6] enabling systematic task decomposition. TSL is a sequence of user actions grouped into 6 interaction steps that capture both the decision-making actions as well as the physical actions. The 6 steps are (1) Identify Task/Procedure, (2) Select Function, (3) Access Function, (4) Enter data for Function, (5) Confirm, and (6) Monitor. Predictions are based on the salience of the cue to prompt the next operator action [6]. The diagram in Figure 2 below is a generic TSL example showing the interactions within the cockpit where the agents involved in the procedure performance are represented by lifelines and can be (1) a number of devices including the device to be certified, (2) one or two operators (pilots) depending on the task, and (3) the outside world. The TSL steps are shown in red.

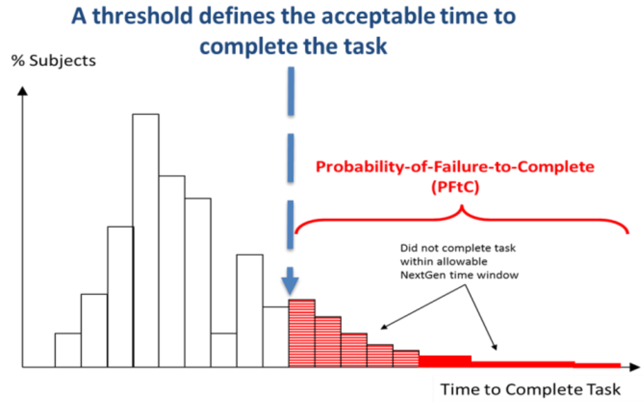

FIGURE 1
TIME TO COMPLETE PROCEDURE AND PROBABILITY OF FAILURE TO COMPLETE

The arrows represent the information circulating between the agents including the operator's cognitive actions represented by recursive arrows. The arrows progress in a sequential fashion from top to bottom to show time progression. The "Identify task/procedure", is first triggered by semantic cues from the device or/and the outside world, followed by a memory item retrieval leading to the "Select Function" guiding the operator to look in the relevant area of the cockpit/device. One or multiple (k) visual cues are presented to the operator which in turn triggers the "Access" step. The operator then "Enters" the required data causing a display change in the device represented by a semantic cue prompting the "Confirm", after which semantic cues either from the device, or the outside world, or both enable the "Monitor" step. The TSL steps are expanded and modeled as Operator Actions (OA) in the PAT.

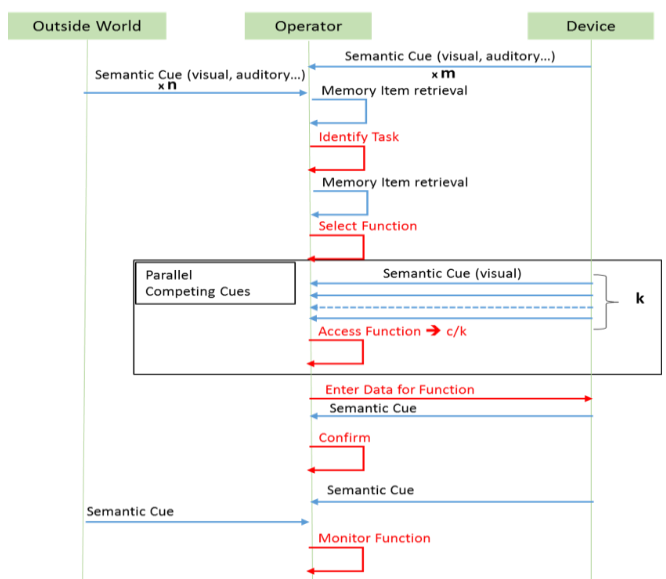

FIGURE 2
TSL FRAMEWORK REPRESENTED IN AN INTERACTION DIAGRAM

The OAs are then categorized into pre-defined Operator Action Categories (OAC) and assigned to a statistical distribution of time to complete the operation. After that, the compilation of all OA times is performed to compute the total procedure time.

STAKEHOLDERS ANALYSIS

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" [4]. 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 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, conflict of interest and tensions may arise between these two stakeholders leading to pressure on the DER.

PROBLEM STATEMENT

The new regulation FAR25.1302 requirement to evaluate all Procedures is cost and time prohibitive if it has to be accomplished by the currently approved 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, and the subjective nature of evaluation.

PROPOSED SOLUTION

I. CONOPS

The proposed solution is a systems engineering tool for use by the DER that substitutes the current certification evaluation process (i.e. Inspection, and HitL on all tasks) by Inspection accompanied by PAT testing on all tasks to dedicate HitL testing to only procedures with high PFtoC. The PAT simulates flightcrew performance and keeps track of their timing while executing procedures on a device enabling affordable automation for task/usability analysis. The automated evaluation is significantly faster and cheaper. It also eliminates stakeholders' tensions by providing rapid, yet objective means of analysis. Also, in addition to enabling compliance with the requirement to analyze all tasks, it enables implementation of accessible and usable control and information by the flight crew in a manner consistent with (1) urgency; as a result of defining a threshold for the PFtoC, (2) frequency; as a result of modeling OACs with various frequency levels (rare, moderate, and routine), and (3) task duration; as represented by the TTC the procedure.

A high level idea of the PAT concept of operations is illustrated in the Figure 3 below where the DER evaluates a proposed system design (device) and its allotted airline SOP using a combination of TSL and operator performance model that includes a database of OAC and their related statistical distributions to generate time-on-procedure distributions for each procedure.

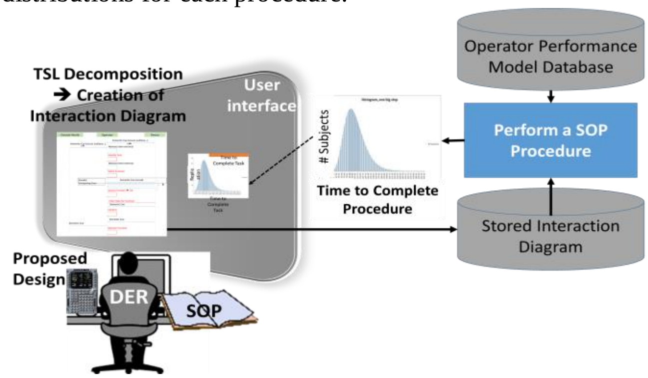

FIGURE 3
PAT CONCEPT OF OPERATION

II. Mission Requirements

The main PAT mission requirements are:

MR.1 The PAT shall enable compliance with the FAR 25.1302 requirements.

MR.2 The PAT shall reduce the post FAR25.1302 certification cost by at least 65%.

MR.3 The PAT shall remove the stakeholder tensions between the FAA, the Applicant, and the DER.

III. Derived Requirements

The main Technology Requirements are as follows:

TR1. The PAT shall use a Graphical User Interface to enable user input and tool output display.

TR2. The PAT shall use a Database to store Operator Performance data.

TR3. The PAT shall accept up to 600 procedures for decomposition.

TR4. The PAT shall use a lookup function to map the OA to their statistical distributions.

TR5. The PAT shall use an algorithm to generate random numbers and sum them into total procedure time.

TR6. The PAT shall use Monte Carlo Simulation to simulate 500 operators performing the SOP procedure.

METHOD OF ANALYSIS

I. Overview:

The PAT models perceptual, cognitive, and motor skills time performances to simulate pilot operations in the cockpit environment. The input to the PAT is a detailed description of the tasks from the SOP and/or the CBT. The PAT combines SOPs with a database of Operator Performance Model expressed in statistical distributions of time to complete an action in seconds. The output of the tool is a frequency distribution of the number of pilots plotted against the time to complete the task along with the PFtoC the procedure. Figure 4 shows a model for the tool with input, output, and where the Operator Performance Model database comes into play.

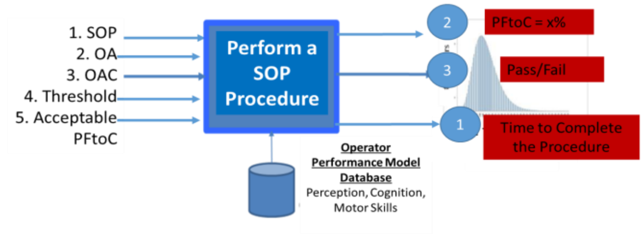

FIGURE 4
PROCESS MODELED BY PAT

The process is seven steps. The DER performs steps 1 and 5 while the rest is accomplished by the PAT as follows:

- Step1. Decompose procedure into Operator Actions (OA) and categorize into OAC.
- Step2. Associate with Operator Performance Database (OPD).
- Step3. Simulate task operation.
- Step4. Request threshold for allowable time window and acceptable PFtoC from user.
- Step5. Provide user input as required.
- Step6. Calculate PFtoC for the procedure to determine if device functions require further HitL testing.
- Step7. Perform a Sensitivity Analysis to identify OAC with highest impact.

Step one is performed using an interaction diagram (see Figure 6). Each OA falls under an Operator Action Category (OAC) associated with a statistical distribution as shown in Table I. After the identification of the OAC for each OA, random numbers are generated according to their statistical distributions before being aggregated to a procedure time. This operation is performed 500 times using a Monte Carlo simulation to result in a distribution as shown in Figure1.

II. Mathematical Models/Algorithms

1) OAC Statistical Distributions Elicitation

There are 17 OACs in total (shown in Table I). Each OAC is accompanied by a statistical distribution. The statistical distributions and their parameters were either taken from literature review (ex: Visual Cue N~ (0.62, 0.11) [8]), or produced by designing experiments using GMU students as subjects. For example, the "Working Memory Item Retrieval" OAC was performed between a time keeper and the subject as follows:

1. Time keeper reads out loud all of the steps of a task. Example the "Hold at Present Position" Task
2. Time keeper immediately starts stopwatch for 50 seconds allowing subject to remember what was read
3. Time keeper asks question related to the above mentioned task, and immediately starts stopwatch until subject answers correctly. The recorded time is labeled T_{tot}. Note: The timer is kept running if the subject answers the question incorrectly.
4. Once the correct answer is obtained, subject speaks out the answer again while time keeper records timing of voicing out the answer. This time is labeled T_{ans}
5. Repeat steps 2 to 4 five times using a new question. For each question, compute the working memory item retrieval time T_{WM} by subtracting the answer time from the total time as follows:

$$T_{WM} = T_{tot} - T_{ans} \quad (1)$$

6. Record and analyze data for fit.

This experiment was performed by 5 subjects for 2 tasks with 5 questions each resulting in 50 data points. In this case, the distribution was found to fit a lognormal type according to the analysis [9] of the mean, variance, and skewness of the sample. The resulting distribution is Log (0.737, 1.21).

TABLE I
OPERATOR ACTION CATEGORIES AND THEIR STATISTICAL DISTRIBUTIONS

Operator Action Category (OAC)	*Distribution*
Button push	Normal(0.3, 0.01)
TSL procedure identification	Normal(0.5,0.002)
Decision/choice of button	Lognormal(2,1)
flight controls manipulation with feet	Normal(2.31, 0.651)
Flight controls manipulation with hands	Lognormal(1.12, 0.231)
Listen to audio (ATC Clearance, Aural Warning)	exponential (1)
Long term memory item retrieval	Lognormal(1.21, 2.1)
MCC callout	Normal(2,1)
MCC readback	Normal(1.5, 0.2)
Talk to ATC	exponential (0.9)
Thrust levers manipulation	Triangular(0.1,2,3.5)
TSL Function Selection	Normal(0.1,0.002)
Visual check	Normal (1, 0.5)
Routine visual cue	Normal(0.632, 0.011)
Moderate frequency visual cue	Normal(1, 0.02)
Rare frequency visual cue	Normal(1.632, 0.03)
Working memory item retrieval	Lognormal(0.737, 1.21)

2) Random Number Generation

The generation of the random number per OAC statistical distribution was performed based on the inverse function and the Linear Congruential (LCG) methods [10].

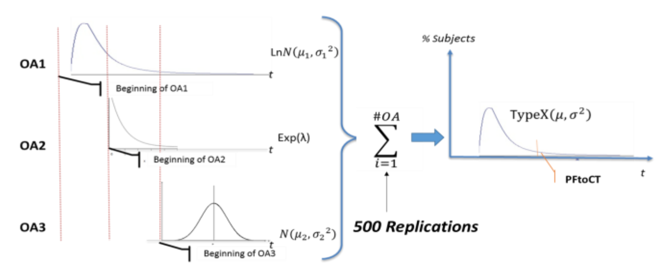

FIGURE 5
SEQUENTIAL VARIABLES PROCESS

3) Sequential Variables Process

The OAs are modeled in series and their sum is illustrated by figure 5 demonstrating a simplified version of a task composed of 3 OAs.

The principles mentioned above are implemented in the below algorithm as follow:

```
carry out any initializations required.
while i < 500
{
      initialize Procedure time PT
      user input ← OA and OAC
read in the OAC for the OA in sequence.
generate next random number as per OAC.
add the number to the accumulated sum PT.
      i++
}
 identify the type of distribution.
 evaluate the average.
 evaluate the standard deviation.
    user input ← Threshold of probability of
failure to complete.
 calculate the probability of failure to
complete the task.
```

III. Case-Study of 15 SOP Procedures

For the preliminary analysis, a pool of 15 procedures was defined for the MCDU. In an effort to make this first sample a representative one, procedures were drawn from across all phases of flight. For example, the "Initialize Position" procedure would be performed during the Pushback/Taxi/Take-off phase. The selection of procedures per phase of flight is as follows: 4 Procedures for Flight Planning, 3 for each of the Pushback/Taxi/Take-off, and Domestic/Oceanic Cruise, 4 for Descent/Final Approach phases, and 1 Procedure 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 6. This procedure totaled 55 operator actions. The OACs weights were as follows: (47.27%) "Routine visual cue", (16.36%) "Long-term memory item retrieval", (12.73%) "Decision/Choice of button", (10.91%) "Button push" and (5.45%) "Working memory item retrieval" while (1.82%) of the procedure is either "TSL Procedure Identification", "Listen to Audio", "TSL Selection", or "Visual Check".

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

RESULTS

The PAT ran 500 replications for each of the 15 procedures analyzed. Table II below summarizes the PFtoC. For demonstration purposes (and lack of user input), the threshold for PFtoC was set to 3 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, and were therefore flagged for HitL testing.

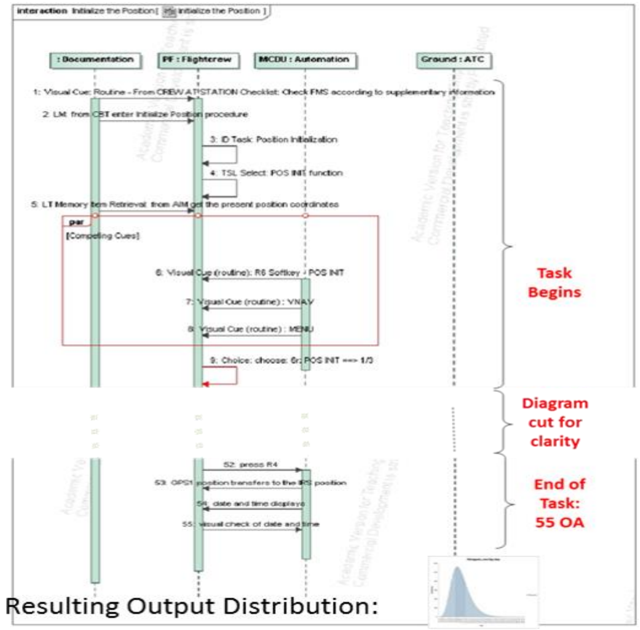

FIGURE 6
INITIALIZE POSITION PROCEDURE DECOMPOSITION EXAMPLE

Evaluating the 15 procedures with PAT took 12:47 hr. entry time, 11.56 sec. average runtime, and analysis showed that PFtoC ranged between 0.4% and 2.6%. Also, 3 over 15 procedures (20%) were flagged for HitL testing. The *Visual Cue* category gathered the greatest OAC percentage which highlights the importance of semantic cues to recognizing emerging mission situations and prompting the next step of the procedure process.

TABLE II
SUMMARY OF RESULTS

#	***Procedure***	***PAT Entry Time [hr.]***	***PAT Run Time [sec]***	***Threshold u+ 3std [sec]***	***PFtoC (%)***	***OAC Visual Cue Count (%)***
1	Initialize Position	1:05	3.76	814.52	1.60	47.27
2	Enter Company FP*	1:30	3.21	742.43	2.60	40.82
4	Check FP* Progress	1:00	3.66	1200.25	1.00	45.16
5	Switch MCDU ON	0:15	19	386.05	3.40	41.18
6	Enter Runway info	0:30	16	578.11	1.00	59.57
7	Enter FP* manually	1:28	13	733.28	2.60	52.00
8	Hold Present Position	2:30	10	814.52	1.50	39.00
9	Change turn …	0:21	14	533.30	1.60	39.29
10	Check Climb/Cruise…	0:23	13	418.68	1.80	48.00
11	Change Speed…	0:36	13	274.94	2.20	51.51

12	Select Arrival…	0:47	14	538.91	2.20	61.70
13	Enter a Runway …	0:36	13	385.47	2.40	41.18
14	Enter Hold Exit…	0:26	13	708.35	2.40	50.00
15	Select Speed…	0:40	11	539.50	0.4	48.0
--	**Total Time**	**12:47**	**173**		--	--
--	**Average**	0:47	11.56		2	47

*FP: Flightplan

VALIDATION

Two sets of validations took place: the first set is the validation of the OAC statistical distribution against literature review, and the second employs experiments using GMU students performing tasks on a device.

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: an (1) "A La Carte" model priced at $200 per function, a (2) License Seat model priced at $75k per license in addition to a yearly fee, and a (3) an Application Purchase model priced at $500k per unit in addition to a yearly recurring fee. The market analysis identified 10 potential buyers among which Honeywell, Garmin, Boeing, Rockwell/Collins, Avidyne, Smiths, G.E. etc. are listed. A business model was built under the assumption of a gradual market penetration with a shift from the *A la Carte* model to *the Application Purchase* model as depicted in the chart in Figure7 below.

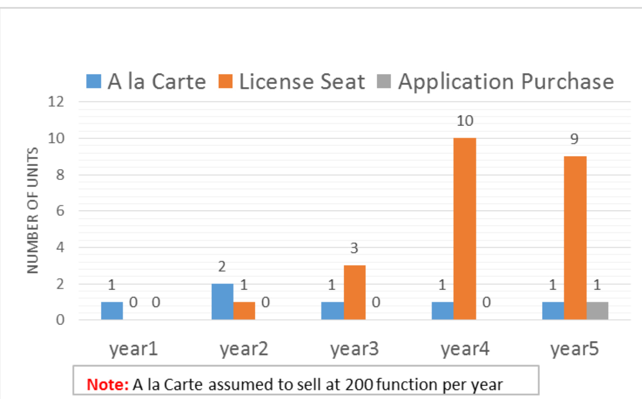

FIGURE 7
YEARLY SALES PER BUSINESS MODEL

The PAT's fixed costs are $250k in software development, and $250k in marketing, sales, and business management. A recurring cost is appraised to start at $250k and to scale with consulting business and support services growth. The cumulated cost, revenue, and profits along with the Return on Investment (ROI) were calculated over a five years period. Results are displayed in the table below:

TABLE III
COST, REVENUE, AND PROFIT STREAMS

Time	*Cost*	*Cumulated Cost*	*ROI*
year0	500000	500000	-200%
year1	250000	750000	-193%
year2	250000	1000000	-165%
year3	250000	1250000	-130%
year4	250000	1500000	-30%
year5	250000	1750000	20%

The ROI is 20% at the end of year 5. The breakeven point is forecasted to happen after 4 years as displayed on the boxed rows in Table III above.
An enhanced version of the business model including a demand queuing model and a worse and best case scenarios over a 10 years' timeframe is planned for the presentation.

CONCLUSIONS AND RECOMMENDATIONS

Analysis of the human factors certification process shows that the PAT reduces the evaluation time by 68% per function in addition to enabling evaluation of all tasks. Therefore, the PAT solves the problem imposed by FAR25.1302. Recommendations include further work on understanding the semantic cues leading to the recognition of emerging mission situations and the next step of the procedure process.

ACKNOWLEDGMENT

Mr. George Blaha, Raytheon. Dr. Lance Sherry, GMU.

REFERENCES

[1] Graeber, R.C.; Emmerson, P.; "Human Factors – HWG Final Report," FAA, 2004.
[2] U.S. Government Publishing Office, "GPO," COOP, 23 March 2015. [Online]. Available: http://www.ecfr.gov/cgi-bin/text-idx?SID=dcb751a5dec956359bc6dbf9db0457b0&node=se14.1.25_11302&rgn=div8. [Accessed 25 March 2015].
[3] Department of Transportation, "Installed Systems and Equipment for Use by the Flightcrew," Federal Register, Proposed Rules, vol. 76, pp. 6088-6094, 2011.
[4] FAA, "Advisory Circular 25.1302," 2013.
[5] L. Sherry and M. Feary, "Technique for Mean of Compliance (MOC) for FAR 25.1302 for New Flightdeck Technologies," FAA Human Factor Research, vol. Annex 4, no. 5, 2011.
[6] L. Sherry and e. al., "Estimating the Benefits of Human Factors Engineering in NextGen Development: Towards a Formal Definition of Pilot Proficiency," in 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO), 2009.
[7] E. Patton and W. Grey, "SANLab-CM: a Tool for Incorporating Stochastic Operations into Activity Network Modeling," Behavior Research Methods, vol. 42, no. 3, pp. 877-883, 2010.
[8] S. Cao, "Queueing Network Modeling of Human Performance in Complex Cognitive Multi-task Scenarios," 2013.
[9] E. S. W. A. M. Limpert, 1 May 2001. [Online]. Available: stat.ethz.ch/~stahel/lognormal/bioscience.pdf. [Accessed 27 March 2015].
[10] J. Banks, J. S. Carson II, B. L. Nelson and D. M. Nicol, Discrete-Event system Simulation, Prentice Hall, 2009.