2007

Beyond Electronic Flight Bag (EFB) Approval: Improving Crew Performance

Thomas L. Seamster Ph.D.

Barbara G. Kanki Ph.D.

BEYOND ELECTRONIC FLIGHT BAG (EFB) APPROVAL: IMPROVING CREW PERFORMANCE

Thomas L. Seamster, Ph.D.
Cognitive & Human Factors
Santa Fe, New Mexico

Barbara G. Kanki, Ph.D.
NASA Ames Research Center
Moffett Field, California

As operators evaluate and implement Electronic Flight Bags (EFBs), the emphasis has been on their operational approval and certification. This research provides data that demonstrate how an operator can aim beyond the limited objectives of the EFB approval process to improving crew performance. This paper reports on evaluation results that show how crews working with an EFB can not only equal, but can exceed the performance of those working with traditional paper documents.

Introduction

The Federal Aviation Administration (FAA, 2003) outlines the EFB certification and approval process in Advisory Circular 120-76A. The thrust of the Advisor Circular is that an EFB needs to be as good as the existing paper system, and that it not result in "unacceptable flightcrew workload." This guidance extends to the mounting system, the applications and to its overall use and stowage. We report on results that show how crews working with the EFB can not only equal, but can exceed the performance of those working with traditional paper documents.

This is part of a full EFB evaluation that included an operational evaluation conducted in full-motion simulators. This paper presents data collected during that operational evaluation that have been analyzed to determine specific points during ground and flight operations where crew performance can be improved. The iterative and operational nature of this EFB evaluation goes beyond checking the EFB against a list of human factors considerations to the identification of ways for enhancing crew taxi awareness as well as crew Flight Management System (FMS) data cross checking. The objective was not just to seek operational approval for a piece of hardware and its software, but to implement an entire system that includes the EFB with its operational philosophy, procedures, training as well as standard crew assessment in ways to measurably improve crew performance.

The operational evaluation included an experiment designed to compare pilot workload, head-down time, performance and crew feedback with one set of crews working with traditional paper flight deck documents and a second set working with electronic documents. Forty volunteers were solicited from within a single fleet of the operator's pilot population. Twenty volunteers were Captains and the remaining 20 were First Officers (FO). The volunteers were formed into 20 crews, each made up of a Captain and FO, randomly assigned to either the Electronic or Paper condition. Instructor/Evaluators (I/Es) were then assigned to crews based on their schedule ensuring that each I/E conducted at least two Electronic and two Paper Line Operational Evaluation (LOE) sessions.

Instructor/evaluators rated crew performance and head-down time, while the pilots rated workload for each phase of the flight. The results show specific areas where EFB use improved performance and reduced workload and head-down time compared to the traditional paper documents. The analysis also indicates areas where performance, workload and head-down time could be further enhanced. The paper concludes with recommendations on how operators can help their crews, through the EFB implementation process, improve performance while reducing their workload.

Electronic Flight Bag

The EFB must be evaluated as a complete system including the hardware, software, flightdeck procedures and training. EFBs cover a wide range of hardware and software that have been further classified by the FAA. Hardware is divided into three classes: Class 1 hardware are portable computing devices that are not mounted to the aircraft. Class 2 EFBs are computing devices that are attached to the aircraft during normal operations while Class 3 EFBs are installed on the aircraft allowing for a wider range of applications. The three software types allow for greater computation and complexity as they move from Type A through to Type C software. The majority of Type A applications are static manuals, data, and tables. Type B software includes a range of dynamic information including navigation charts and weight and balance calculations. Type C applications include the primary flight displays. Each Class of hardware and each software Type may have different approval requirements with numerous permutations as one considers the different possible hardware and software combinations.

EFB Hardware

For this evaluation, the EFB hardware was classified as a Class 2 device consisting of a central processing unit (CPU), a display and an EFB mount. The CPU included the computer, a lithium ion backup battery, USB and monitor connectors, and the On/Off switch. The 1.6 GHz Centrino Mobile CPU was used for this evaluation with 512 MB of random access memory and a 40 GB hard drive.

The display consisted of a 10.4 inch touch screen with standby and brightness controls and hardwired cable for connection to the CPU. Other than turning the system on and off, all other flight deck EFB functions were initiated through the touch screen. The resistive touch screen was a 10.4 inch color TFT LCD with a 1024 x 768 resolution when used in the landscape mode as evaluated. The film-on-glass touch screen was engineered for direct sunlight readability with an illumination from 3 to 750 nits. The overall size of the touch screen with bezel was 7.4 inches wide by 11 inches high and six tenths of an inch thick. The approximate location of the Captain's touch screen is shown in Figure 1.

EFB Software

The EFB software was classified at Type A and Type B. The applications being evaluated operated under Windows XP and included a full set of charts, operating manuals, airport operational information and the Minimum Equipment List (MEL). The charts included airport ground, low visibility, airport parking, airport facility, standard instrument departure, standard arrival route and en-route charts.

EFB Procedures and Training

The operator had developed extensive EFB Standard Operating Procedures (SOPs) that reflected their automation philosophy. General EFB procedures were developed for each phase of flight, and specific SOPs were developed for EFB initialization, its use in the takeoff briefing, approach planning, and in case of EFB failure. The crew training consisted of a three hour home study session followed by a two hour classroom session that included an introduction to the EFB, its applications, its SOP as well as a set of navigation questions and problems (for more detail on the training see Kanki & Seamster, 2007).

Figure 1. Approximate location and configuration of the touch screen during EFB testing

Methods

The operational evaluation was designed as an experiment that compared crew performance, pilot workload, head-down time and crew feedback based on one set of crews working with paper documents and a second set of crews working with electronic documents.

Participants

Forty volunteer pilots were solicited from the operator's fleet. Twenty volunteers were Captains and the remaining 20 were First Officers (FO). Participants had to be available within a four week period for a six hour late night EFB evaluation session. The volunteers were formed into 20 crews each made up of a Captain and FO and randomly assigned to either the Electronic or Paper document condition. Instructor/evaluators were then assigned to crews based on their schedule ensuring that each I/E conducted at least two Electronic and two Paper sessions. Pilots had an average of 1,359 hours on fleet type with a range from 96 to 6,000 hours. All participants were experienced pilots with an average of 7,458 total flight hours with a range from 500 to 15,000 hours.

Scenario

A Line Operational Evaluation (LOE) scenario was developed to fully exercise the use of flight deck documents with an emphasis on the evaluation of how the documents were used by the crews with no additional training taking place during the simulator session. The scenario was a three hour flight from A to B with a divert due to runway problems at B. This scenario was made up of six event sets that required operating document usage during Preflight, Engine Start, Cruise with several route changes, a divert, and a low visibility Taxi-In. Each event set was defined by a specific period of time designed so that the I/E could attend to the evaluation of each crew's use of documents, time to complete specific tasks, and the degree of head-down time of each crew member.

Prior to the LOE session, crews were briefed that it was the Captain's leg and that they should treat this as a line flight simulation including use of headphones, seatbelts/shoulder harnesses and Standard Operating Procedures (SOP). In the EFB condition, crews were told to use EFB SOP, adjust the display for optimal viewing, and use the EFB to access their charts, Flight Operational Manual (FOM) and Minimum Equipment List (MEL).

Design and Procedure

The operational evaluation was designed to determine how a range of crews interact with paper or electronic documents across all phases of flight. The primary design objective was to assess individual pilot workload and the resulting pilot performance as they interacted with their operating documents. Additional crew comments were collected following the simulator sessions to elicit what the pilots thought worked best with the EFB and what needed additional work.

Each of the 20 crews worked with one I/E throughout the 6 hour session consisting of a briefing, an LOE and a debriefing. Each crew came to the session having completed a two and a half hour EFB home study unit develop to familiarize pilots with the new electronic charts and the software used to display those charts. The briefings were held in a training center briefing room with the I/E introducing the crew to the EFB operational evaluation and the purpose of informed consent. Crews were asked if they had any questions, and they then read and signed the Informed Consent Form.

The briefing then continued with demonstrations and practice on the EFB and the new electronic navigation charts**.** During this EFB and charts briefing, crews worked on individual desktop computers with a high fidelity simulation of the EFB software. The I/E walked them through a familiarization training session that lasted approximately 100 minutes. The briefing included a short break and EFB practice session. After a brief quiz, crews were told the condition to which they were randomly assigned, either Paper or Electronic. They were then given the flight plan and asked to do their flight planning and take a break before the LOE.

For the LOE, crews under the Paper condition used their Jeppesen black and white charts and paper documents that they had been working with on line flights. Crews under the Electronic condition used the EFB working with the a new set of color charts and other electronic documents including the MEL and the Flight Operations Manual (FOM). Both conditions used the paper version of the QRH. Crews met the I/E outside the simulator and were given a LOE orientation and then were instructed to proceed as if they were conducting a line flight. Two identical full motion flight simulators were used for the LOE sessions. Following the three hour LOE, while still in the simulator, crews completed the Workload Rating Form based on the NASA Task Load Index (TLX) (Hart & Staveland, 1988), and then they took a short break.

Finally, crews met with the I/E in the debriefing room where they each completed an exit questionnaire and participated in the final debrief. Crews were given an opportunity to ask any questions and were thanked for their participation.

Instructor/Evaluator Standardization

Since the I/Es were to provide a substantial amount of the EFB operational evaluation data, extensive rater training and practice were developed to ensure an acceptable level of standardization. Four I/Es were trained and evaluated using specific behavioral markers to rate each item in each event set based on the Inter-Rater Reliability (IRR) outlined in Schultz, Seamster, and Edens (1997) and Williams, Holt, and Boehm-Davis (1997). The I/Es were further trained and evaluated in rating head-down time.

Prior to conducting the EFB operational evaluation, I/Es were trained on a 6-point scale to rate pilot performance. They then rated pilot performance by viewing a video recording of a crew flying the six event sets working with an EFB. That rating data were analyzed to determine I/E standardization. Instructor/evaluator head-down ratings were analyzed using the same standardization training and evaluation methods.

The I/E rating training took place over three different days. The first four modules in the I/E the training provided an understanding of usability testing, operational approval of the EFB, the design and rating of the LOE event sets and an overview of rating pilot performance. The remaining modules helped to develop the knowledge and skills required to assess pilot performance and head-down by event set and to standardize the assessment process.

The Final Standardization Check was the IRR module following the principles outlined in Holt, Hansberger, & Boehm-Davis (2002). Reliability was computed based on I/E ratings of a single taped crew performance of the LOE recorded in a full-motion simulator flying the LOE session used for this operational evaluation.

The Final Standardization Check data were analyzed to determine the degree of agreement between the four I/Es who would be evaluating the LOE sessions. Systematic differences for crew performance ratings were computed by comparing the individual mean ratings with the group means, and none of the I/Es were significantly more lenient or harsher than the group. As a further check, I/E ratings of the Final Check Video were combined for each event set and subjected to a Multivariate Analysis of Variance (MANOVA). Only Event Set 3, the change in clearance during climb, showed a significant difference between I/Es. The Tukey-Kramer Multiple-Comparison Test indicated that the mean ratings for one I/E were significantly higher than that of two others. Although the IRR results showed no systematic differences, there were some indications that one of the four I/Es rated more leniently than two others or, conversely, two of the I/Es rated significantly more harshly. This pattern is further investigated in the Discussion section.

Results

The results reported in this section include crew performance, head-down time and workload.

Crew Performance

During the LOE, I/Es rated crew performance with flight deck publications for each of six event sets. Individual pilot performance was rated as each crew member performed specific tasks and interacted with the electronic or paper documents. Each evaluation was made within the appropriate event set covering preflight, engine start, climb, two route changes and taxi-in. Instructor/evaluators used a 6-point scale to make their crew performance evaluations where:

6 = Outstanding,
5 = Superior
4 = Good
3 = Minor Problem
2 = Minimally Acceptable
1 = Unacceptable.

A MANOVA was conducted to determine the effects of the Electronic vs. Paper condition across all ratings of performance for each crew member. Of the 14 ratings conducted during the LOE, nine were conducted for both the Captain and First Officer. All nine items showed superior evaluations for crew members using Electronic documents with on average a half rating point improvement over crews using the traditional paper documents. The ratings for Event Set 3 (FMS), Event Set 4 (SOP), and Event Set 6 (SOP) were significantly higher for crew members working with Electronic documents as shown in Figure 2. Event Set 1 and 2 were during preflight and pushback when crews had their first exposure to the EFB hardware and software. By Event Set 3, they had spent about 45 minutes with the EFB and had a greater level of familiarity with the new system.

Head-Down Time

Head-Down Time was rated by the I/E for each crew member during each of the six event sets. A MANOVA was conducted to determine the effects of Condition (Electronic vs. Paper) across all ratings of Head-Down Time for each crew member. Crews under the Electronic condition had significantly shorter Head-Down Times in Event Sets 3, 4, 5, and 6 as shown in Figure 3.

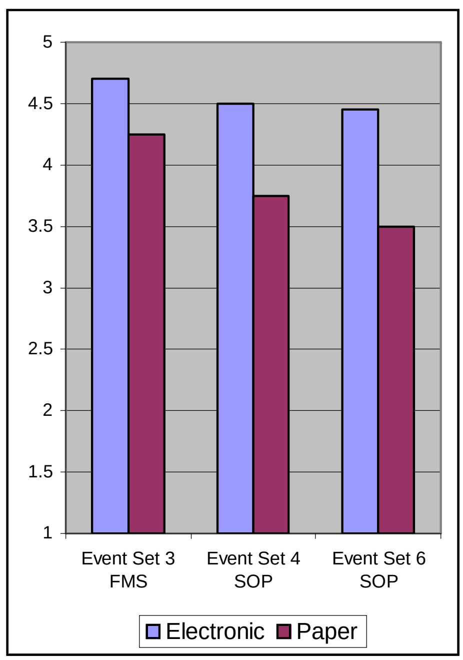

Figure 2. Significantly Different Crew Performance ratings for Electronic versus Paper Documents by Event Set (1 = Unacceptable and 6 = Outstanding).

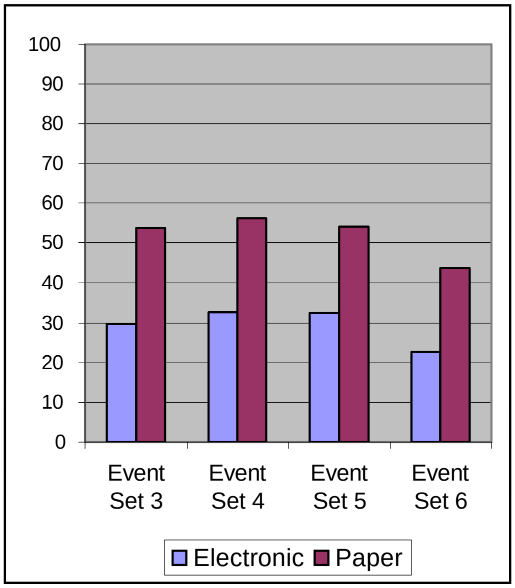

Figure 3. Significantly Different Crew Head Down-Time ratings for Electronic versus Paper by Event Set (0 = Minimal and 100 = Excessive)

Head-Down Time for individual crew members working with Electronic documents was less in each event set by an overall average of 18.8 points on a 100 point scale. The mid-point on this rating scale, 50, is equivalent to the Head-Down Time for an average crew during that phase of flight, and in all cases (see Figure 3), the average for crews under the Electronic condition was well below that average Head-Down Time.

Workload

Pilot workload while flying with either the electronic or paper documents was rated by each pilot immediately following the LOE. A NASA TLX-based questionnaire was administered for each event set in which pilots were asked to make individual ratings on six dimensions including: 1) Mental demand, 2) Physical demand, 3) Rushed, 4) Degree of success, 5) Degree of work and 6) Stress or annoyance.

A MANOVA was conducted to determine the effects of the Electronic vs. Paper condition across all ratings of workload by each crew member. Twenty six of the 36 items rated showed no significant difference between pilots using electronic or paper documents. The ten items that were significantly different are shown in Figures 4 and 5. All but one of those items, Event Set 1 "How rushed was the pace of the Preflight?," showed significantly lower workload or higher success for Electronic compared with Paper.

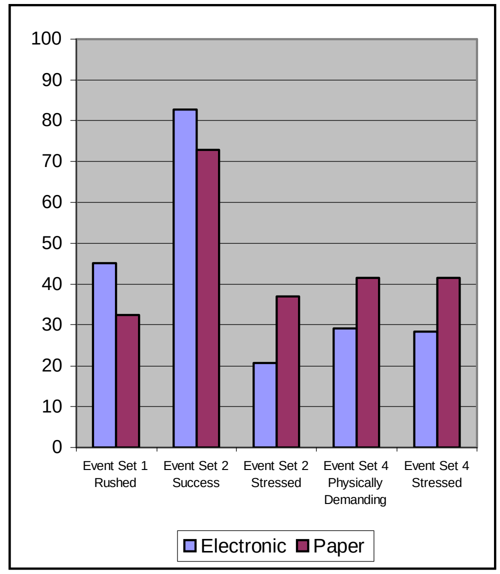

Figure 4. Significantly Different Crew Workload ratings for Event Sets 1 through 4 (Electronic versus Paper) where 0 = Very Low and 100 = Very High

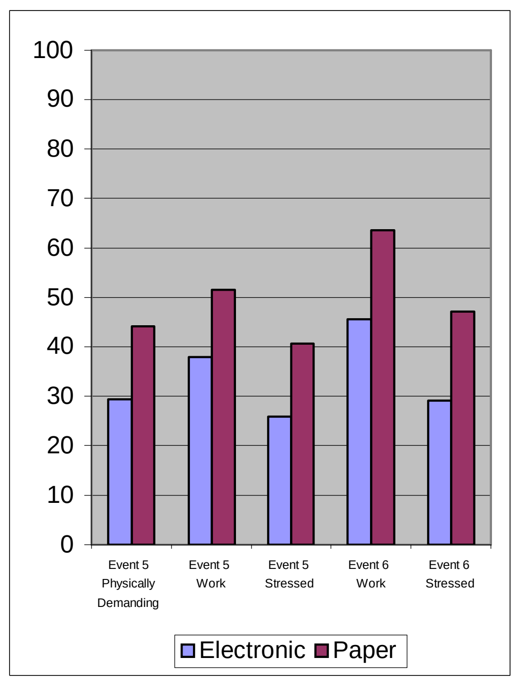

Figure 5. Significantly Different Crew Workload ratings for Event Sets 5 and 6 (Electronic versus Paper) where 0 = Very Low and 100 = Very High

A review of the significant differences shows that when crews accessed the EFB for the very first time during Event Set 1, flightdeck setup, they felt more rushed than crews working with paper documents. As the LOE progressed, and especially during Event Sets 4 through 6, crews under the Electronic condition felt significantly lower workload along a number of the workload dimensions as compared to crews under Paper. Data from Head-Down Time and Crew Performance also support this pattern of improved EFB usage as the LOE progresses and crews gained greater familiarity with the system.

Discussion

The results combine to demonstrate an EFB induced improvement in crew performance backed by a reduction in head-down time and workload. These results are particularly noteworthy because the crews under the Electronic condition had minimal training and familiarity with the EFB. Further, there was a pattern where performance improvements were more pronounced as crews gained exposure to the EFB during their three hour LOE. This suggests that with good operating procedures, supporting training, and more experience, crews can achieve even more EFB benefits. Crews showed specific improvements when verifying the FMS and in following SOP. With greater experience, crews may not only demonstrate further improvements in these areas but in others such as preflight setup, briefings as well as taxing, especially at airports with complex taxiways.

Precise assessment tools and procedures are essential to obtaining a clear understanding of how EFBs can improve crew performance. This is underscored by the persistence of a pattern where one I/E was more lenient than others starting with I/E training and throughout the operational evaluation. To achieve standard assessments, operators should start working on I/E IRR as early as possible in the transition to an EFB. This provides more time to identify and correct problems in rater standardization, and it also allows operators to collect measures of pilot performance under the existing system before the new one has been implemented. A comparison between pre and post system pilot performance is invaluable in helping operators refine their EFB implementation.

Operators need a comprehensive and usable IRR toolset to ensure reliable measures. As Holt et al. (2002) have documented, reliability is made up of at least five dimensions, and a full set of reliability measures can provide I/Es with a clear picture of their standardization. An IRR toolset is needed that provides complete reliability results without overloading the I/Es with too much data. There is also a need for a toolset that allows the operator to compute ongoing IRR based on I/E ratings of actual crew performance during training sessions and line checks. If operators implement the EFB as a complete system to include new procedures, targeted training, and precise measures, they should realize marked improvements in not only EFB and operating document related tasks, but also in overall crew and SOP performance.

Acknowledgments

We acknowledge the contribution made by the NASA/FAA Operating Documents Group over the past ten years with recognition to Eleana Edens (FAA AFS-230) for her support of the NASA/FAA Operating Documents Group through FAA ATO-P**.**

References

FAA (2003). *Guidelines for the certification, airworthiness, and operational approval of electronic flight bag computing devices, Advisory Circular AC 120-76A*. Washington, DC: The Federal Aviation Administration.

Hart, S. G., & Staveland, L. E. (1988). Development of a NASA-TLX (Task Load Index): Results of empirical and theoretical research", *in* P.S. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam, The Netherlands: Elsevier Science Publishers.

Holt, R. W., Hansberger, J., & Boehm-Davis, D. A. (2002). Improving Rater Calibration in Aviation: A Case Study. *The International Journal of Aviation Psychology*, *12*(3), 305–330.

Kanki, B. G., & Seamster, T. L. (2007). Optimizing EFB Through Training, Standards, and Best Practices. *Proceedings of the Fourteenth International Symposium on Aviation Psychology*. Dayton, OH: Wright State University.

Schultz, K., Seamster, T. L., & Edens, E. S. (1997). Inter-rater reliability tool development and validation. *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp. 495-499). Columbus, OH: Ohio State University.

Williams, D. M., Holt, R. W., & Boehm-Davis, D. A. (1997). Training for inter-rater reliability: Baseline and benchmarks. *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp. 514-519). Columbus, OH: Ohio State University.