The Automation Design Advisor Tool (ADAT): Development and Validation of a Model-Based Tool to Support Flight Deck Automation Design for NextGen Operations

Angelia Sebok,[1] Christopher Wickens,[1] Nadine Sarter,[2] Stacey Quesada,[1] Connie Socash,[1] and Brian Anthony[2]

1 Alion Science and Technology, MA&D Operation, Boulder, Colorado, USA
2 Department of Industrial and Operations Engineering, Center for Ergonomics, University of Michigan, Ann Arbor, Michigan, USA

Abstract

NextGen aviation will require an even greater reliance on automation than current-day operations. Therefore, systems with problems in human–automation interaction must be identified and resolved early, well before they are introduced into operation. This paper describes a research and software development effort to build a prototype automation design advisor tool (ADAT) for flight deck automation. This tool uses models of human performance to identify perceptual, cognitive, and action-related inefficiencies in the design of flight management systems. Aviation designers can use the tool to evaluate and compare potential flight deck automation designs and to identify potential human–automation interaction concerns. Designers can compare different flight management systems in terms of specific features and their ability to support pilot performance. ADAT provides specific, research-based guidance for resolving problematic design issues. It was specifically designed to be flexible enough for both current-day technologies and revolutionary NextGen designs. © 2012 Wiley Periodicals, Inc.

1. INTRODUCTION

The National Aeronautics and Space Administration (NASA) and the Federal Aviation Administration (FAA) expect that, by 2025, air vehicle operations will increase twofold to threefold compared with 2005. To support this significant increase in air traffic volume, new tasks, procedures, and technologies will likely be introduced to enable NextGen operations (Joint

Planning and Development Office, 2010; NextGen Integration and Implementation Office, 2010). For aircraft to fly safely with reduced separations, satellite navigation will require displays of wake vortex and weather data. In trajectory-based operations, pilots will need to fly along a flight path while meeting time restrictions. In merging, spacing, and very closely spaced parallel approaches, pilots will need to carefully maintain separation from a lead aircraft, while self-separation procedures will require pilots to have accurate air traffic control-like displays (e.g., the cockpit display of traffic information, or CDTI). To support the pilot in handling these new responsibilities, increased automation and radical expansion of existing capabilities will be essential. This means that current-day problems in automation design are going to be an even bigger

concern in NextGen airspace. The continued safety of less-forgiving NextGen operations will depend on the design of automation systems that optimally support pilots in a robust and reliable manner.

This paper describes a NASA-sponsored research effort to develop a model-based tool to support designers in developing and evaluating flight deck automation for future aviation operations. In its current version, the tool focuses on the primary autoflight system on current flight decks, the flight management system (FMS). However, it is also created to provide designers with guidance for applying cognitive engineering principles to radical or "revolutionary" new designs. In this paper, we outline our goals, approach, and methods for developing the automation design advisor tool (ADAT), we describe the tool and its predictions, and we summarize our validation efforts. The end product of this project is a prototype tool developed for NASA to support research- and model-based evaluations of flight deck automation.

2. ADAT DEVELOPMENT

2.1. Goals

Our primary goal in this research effort was to create a tool that supports flight deck designers in creating usable and effective automation systems for future aviation operations. We developed ADAT to provide meaningful, easy-to-understand feedback regarding problems it identifies in the design, and (where possible) to make specific recommendations on how to improve the design. We also provide an educational tool that the designer can use even before beginning a design. ADAT provides access to a broad range of highly relevant literature summaries, as well as including the specific references to the original research articles.

Another goal of the tool was to give designers a better understanding of challenges and issues that affect pilot performance. We accomplish this through the structure of the tool, which includes six modules (display layout, noticing changes, meaningfulness of terms, confusability of graphics and labels, cognitive complexity, and procedures), as well as a testing capability using attention modeling to predict pilot behavior with the system. Figure 1 provides an overview of the ADAT tool structure. The ADAT user enters information on the FMS design of interest. They provide this input to each of the six modules, and the modules evaluate specific design issues that affect pilot performance. Each module provides an overall numeric score (1–10, with 10 being best), rating the design for its quality in terms of the module-specific issue. Once the designer has input FMS design data for the layout and changes modules, the attention models can be run. No additional input is needed for the attention models. The attention models provide the user with a dynamic simulation of ADAT-predicted pilot visual scanning and give the designer salient feedback as to the problems with the display layout and event-driven attention capture. Finally, ADAT also supports designers in making direct comparisons across existing and/or proposed FMS designs. We anticipate that these comparisons will be a key use of the tool, specifically for evaluating individual design features across automated systems.
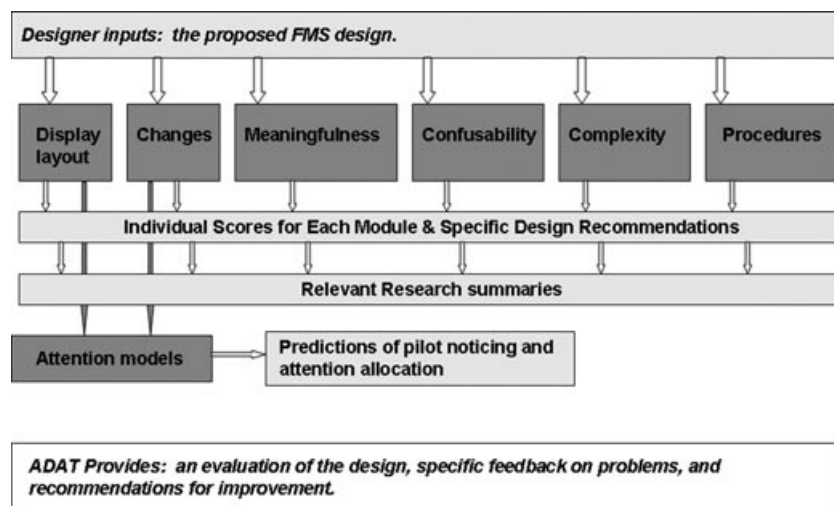
Figure 1   An overview of ADAT, showing the inputs, modules, attention models, and outputs.

This tool is intended to be applicable both to current-day operations with existing flight deck systems and to clean-slate designs for future operations. We achieved this goal by allowing the designer to select and position FMS-related displays and controls on the simulated flight deck with many options. Similarly, ADAT offers the designer significant flexibility in specifying operating modes and procedures. The few assumptions we do make about NextGen operations and associated flight deck designs are that the overall layout of the cockpit will, like current-day designs, include a large out-the-window view, with a pedestal separating two pilot seating areas, and that the primary flight display remains on the flight deck.

The tool is designed to predict the behavior of the "average" pilot. While we recognize that certain design features ADAT penalizes may normally impose only minimal problems for an expert pilot, we also assert 1) that these features may not be trivial under conditions of high stress (Driskell & Salas, 1996), 2) that the pilots most vulnerable to poor design are the inexperienced ones and good design needs to cater particularly to their weaknesses, and 3) that features that improve design usability and clarity will help all pilots.

2.2. Overall Process

Our approach to this work has been an iterative process of researching factors that affect pilot performance, developing rudimentary models, implementing the models in a software tool, and evaluating and refining the models. Throughout the project, we have worked with subject-matter experts (SMEs), including pilots, pilot trainers, FMS designers, and aviation policy and procedures writers. We started the project identifying deficiencies in current-day FMS design, reviewing relevant research, studying operational experiences (incidents and accidents) in which the FMS was implicated, and interviewing SMEs. We then categorized these problems into different design topics.

For each topic or module in ADAT, we specified the information that an FMS designer or other ADAT user would need to enter, given the assumed functionality and interface of the designer's intended design. These inputs allow us to identify potential design issues and problems using algorithms that compute adherence to or departure from established human factors and cognitive engineering design principles. Algorithms were based on both well-known human factors design principles (e.g., layout of information should be based on

frequency of use, sequence of use, importance of the information), as well as on research addressing pilot interaction with the FMS (Fennell, Sherry, Roberts, & Feary, 2006; Funk et al., 1999; Sherry, Feary, Polson & Fennell, 2003; Sherry, Fennell, Feary, & Polson, 2006a, 2006b).

2.3. Problems with the Current FMS

Numerous well-documented problems with the FMS have been identified by several researchers, concerning both its functionality and its interface (Degani, 2004; Hutchins, 1996; Pritchett, 2009; Sarter & Woods, 1992, 1994, 2000; Sherry & Polson, 1999; Vakil, Hansman, & Midkiff, 1995). The classic questions, "What is it doing now?" "Why is it doing that?" and "What will it do next?" are still often heard on modern flight decks (Wiener, 1989). Such questions typically reflect problems with both the functionality and the interface, as well as with the linkages between the two.

Flight deck automation evolved over time, from simple autopilots to the current highly capable and complex FMS. The various autoflight components interact but are not well integrated. As users of these systems, pilots are primarily concerned with, and monitor, the effects of joint autoflight system activities on their current and future flight path. They often have a limited understanding of the system's architecture and logic. Also, the pilots' view of the system is not necessarily consistent with the designers' view, which can lead to misunderstandings and automation surprises (Riley, 2001; Sarter & Woods, 1994, 2000).

Generally, four major classes of problems have emerged through research, field studies, accident and incident reports, and surveys of pilots using the FMS.

1. The elements of the FMS are not integrated but are distributed across the flight deck, requiring considerable scanning and hand movement (Wickens, 2003). For example, the control display unit (CDU), which "programs" the FMS and its associated display, is located quite low in the cockpit, while the mode control panel (MCP), by which FMS parameters can be adjusted, is located quite high. Other relevant displays are distributed between the CDU and MCP.

2. There are well-documented problems that state changes in the FMS are not salient, and hence may not go noticed by pilots,

particularly when these changes are unexpected, occur at high workload, or are triggered by the aircraft crossing some boundary (like a waypoint or altitude level; Nikolic, Orr, & Sarter, 2004; Sarter, Mumaw, & Wickens, 2007; Sarter & Woods, 1992, 1994, 2000).

3. Well-documented aspects of the complexity of the FMS in both its functionality and its interface hinder the pilots' understanding and lead to questions typified by the "What is it doing and why?" (Funk et al., 1999; Wiener, 1989). Many of the complexity issues are compounded by the absence of graphic feedback of the aircraft's future behavior (possibilities and constraints), particularly in the altitude axis (Sherry & Polson, 1999; Vakil et al., 1995).

4. Many complex multistep procedures are required to program the FMS or to modify its instructions in flight in response to an ATC request. The problems in programming are well documented in research and in analysis (e.g., the 21 steps required to reprogram an approach following a controller-requested runway change; Fennell et al., 2006). There is no particular guidance in the interface to lead pilots through the interaction. Pilots need to navigate among pages in the CDU or push buttons before and after turning knobs to implement or make effective changes to flight plans. Of particular concern is that appropriate responses to off-nominal events are typically hidden within the CDU, so pilots—in a stressful situation—are required to remember an infrequently used set of actions to find remotely located information (Fennell et al., 2006; Sherry et al., 2006a, 2006b).

2.4. Modules: Human Factors Principles of Good Design

The following sections will describe in some detail the six modules that the designer uses in ADAT. The modules share the following attributes, which helped to justify their selection and inclusion in the tool:

- Each is associated with at least one identified FMS problem, either related to operational experience, to mishaps, or to research results.
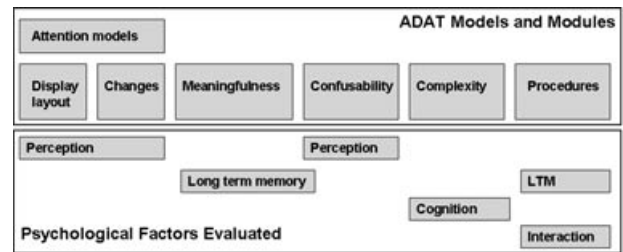

Figure 2 ADAT modules, models, and the psychological factors they evaluate.

- Each module addresses one or more human factors/cognitive engineering guidelines, applicable across a wide range of systems.
- Each module is associated with a body of underlying psychological research, containing either directly applicable computational models or at least model components that were then assembled by the research team to address the module in question. These models serve as the basis for computing module scores by evaluating penalties (for principle violation) or rewards (for principle adherence), that are weighted and summed. Furthermore, to some extent, the separate modules are associated with different domains of psychological research, as shown in Figure 2.
- Each module requires inputs from the FMS designer or other ADAT user that are comparable with the steps that a flight deck designer must go through in creating a design of functionality and interface. This is true whether this design is an evolutionary refinement to a current FMS or a more revolutionary "blank slate" approach.
- Each module provides a design score and, if applicable, identifies specific deficiencies and provides recommendations related to components of the module in question, as described in detail below.

2.4.1. Display Layout

This module is based on well-documented design principles of display layout (summarized by Wickens, Vincow, Schopper, & Lincoln, 1997; see also Ahlstrom & Longo, 2003) regarding location of information based on importance, frequency of use, and sequence of use. The module is based on the fundamental psychological theory of information access effort (Ballard, Hayhoe, & Pelz, 1995; Wickens & McCarley, 2008) and
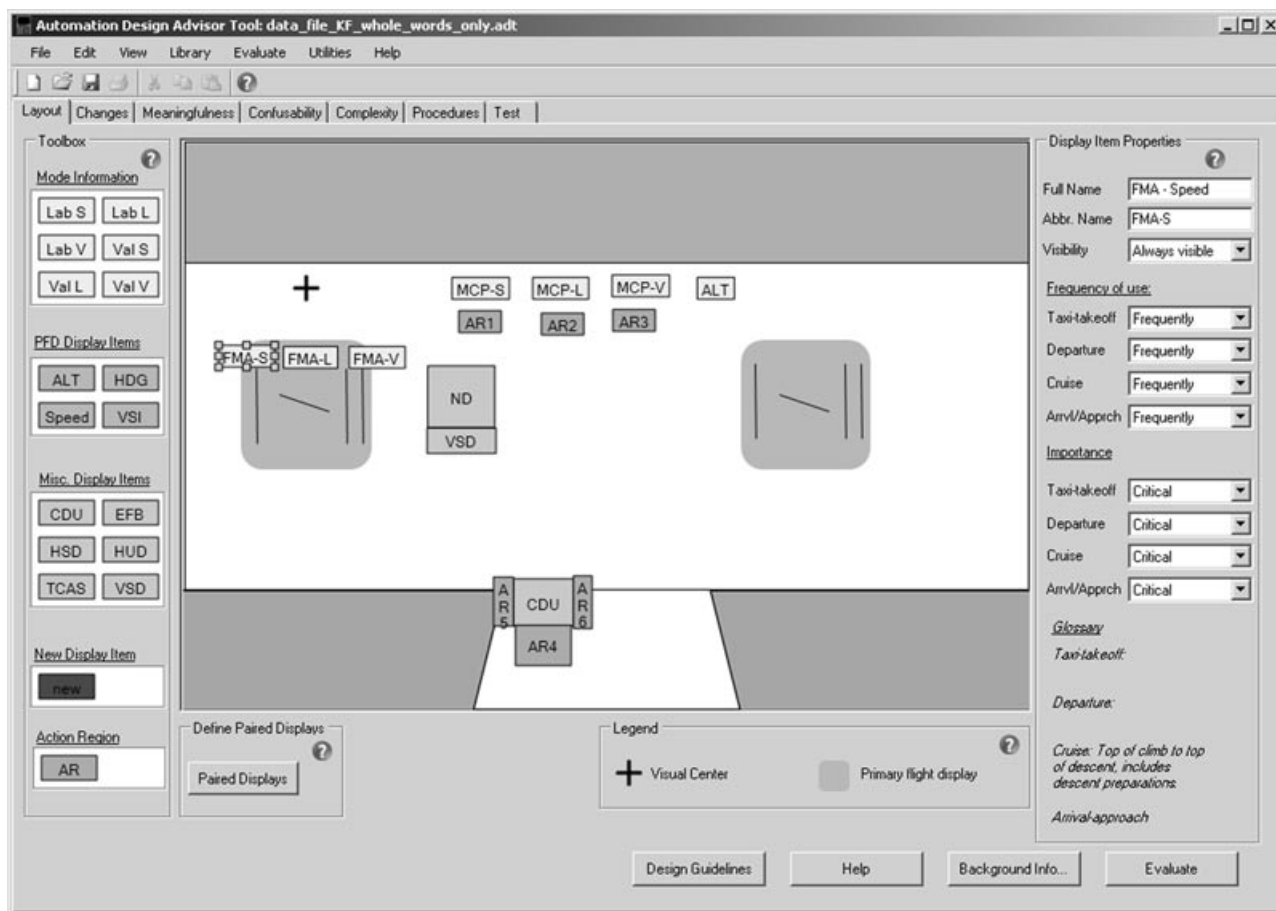
Figure 3 The ADAT layout module, showing the blank slate flight deck. In this figure, the designer has placed several FMS display elements and action regions on the flight deck. FMS elements are identified in toolboxes on the left of the display. The designer clicks and drags these elements to their location on the flight deck. The designer also specifies the properties associated with each display in the area to the right of the figure.

the psychological principle of proximity compatibility (Wickens & Carswell, 1995), which dictates that information that must be mentally integrated should be closely located. The module imposes penalties based on violations of three such principles: displays that are 1) both important and frequently used should be located close to the normal line of sight and 2) always visible; and 3) sequentially used pairs, necessary for information integration, should be located close together.

Because this is typically the first module that the designer encounters, it provides an opportunity to identify critical display and action components and "lay them out" across the flight deck (see Figure 3) so that penalties can be computed. As each display is positioned, its attributes, frequency of use, importance, and sequentially used pair (if applicable) can be specified. Regions where actions (controls) take place are also designated in this module.

2.4.2. Noticing Changes

This module is based on the well-documented finding that pilots often fail to notice the changes in mode status (Nikolic et al., 2004; Sarter et al., 2007; Sarter & Woods, 1994, 2000), even as these changes may signal critically important changes in aircraft behavior. More broadly, it addresses findings from aviation research that pilots have been unaware of critical changes in displayed information (Wickens, 2010; Wickens & Alexander, 2009; Wickens, Hooey, Gore, Sebok, & Koenecke, 2009) that may have major safety implications. For example, in 2006, a midair collision over Brazil (Gol Lineas Aereas Flight 1907 and Embraer EMB-135BJ Legacy 600) was attributed, in part, to the fact that pilots did not notice a rather subtle display change that indicated that their midair collision warning system (TCAS) was no longer broadcasting their position (Fiorino, 2006).

Although human factors principles of alert design (SAE ARP-4102/4, 2007) underlie this module, its psychological basis lies both in the phenomenon of change blindness (Rensink, 2002), as well as factors included in the noticing–salience, effort, expectancy, value (NSEEV) computational model of event noticing (McCarley et al., 2009; Steelman-Allen, McCarley, Wickens, Sebok, & Bzostek, 2009). This module penalizes designs to the extent that important events are signaled by display changes that are not salient, unexpected, or are located toward the pilot's peripheral visual field, particularly when they may be encountered in high workload. A special feature of ADAT is that noticeability of changes can be directly visualized in the attention models (described in Section 2.7).

2.4.3. Meaningfulness of Terms

This module is based on the research findings (Fennell et al., 2006) that otherwise well-trained commercial pilots are not adequately familiar with the meaning of some FMS terms, particularly those used infrequently but of critical importance. The module is derived from findings that designers often choose abbreviations that are not well understood by users (Landauer, 1995). Thus, the meaningfulness module evaluates how understandable terms are by pilots. These include the terms used in the FMS interface: flight mode annunciator (FMA) codes, terms used on the MCP, and terms used in the CDU. Terms are evaluated based on whether they are acronyms, abbreviations, or complete words. Human factors research indicates that words are more understandable than shortened versions and that abbreviations formed by truncating a word are generally more understandable than abbreviations formed by deleting letters in a word (Wickens & Hollands, 2000). In addition, if a single code is used to indicate multiple meanings, the potential for misinterpretation is increased.

A second component of the meaningfulness module is based directly on psychological research on latent semantic analysis (LSA) (Blackmon, Mandalia, Polson, & Kitajima, 2007) that computes meaningfulness of terms based on similarity with other terms. Researchers at the University of Colorado included a corpus of aviation-specific documentation in their LSA capabilities (John, Blackmon, Polson, Fennell, & Teo, 2009), which ADAT references in the meaningfulness module.

In using this module, the designer enters all FMS-related terms whose interpretation will be required by the pilot. The score identifies how meaningful the terms are considered to be, based on both LSA and human factors research regarding term meaningfulness. The meaningfulness of abbreviations, like VNAV, of course, changes with experience, and the current version of the meaningfulness module does not directly address this factor. The module addresses a pilot's understanding through the LSA capability (accounting for part of the overall meaningfulness score), assessing how frequently terms appears in aviation documentation and how related terms are to other aviation terms. This corpus is expected to approximate the commercial aviation vocabulary, so it contributes to the assessment of meaningfulness.

2.4.4. Confusability of Terms and Symbols

The confusability module is designed to avoid problems such as the 1992 Strasbourg crash along the French–German border in which pilots are believed to have confused the setting of descent rate between a flight path angle mode and a feet/minute mode. This confusion, in part, originated from the nearly identical feedback in the two modes and set the airplane on a near uncontrolled descent into the terrain. More generally, it is based on fundamental human factors principles that display entities with substantially different meanings (e.g., ownship symbol vs. traffic symbol, or altitude versus attitude labels) should physically look quite different (Ahlstrom & Longo, 2003; Wickens & Hollands, 2000). For example, the VNAV path and VNAV speed mode labels share much similarity, yet the modes have different behaviors and implications.

The underlying psychological computational model is based on Tversky's (1977) feature overlap model. This is applied separately to the confusability of verbal messages (as, for example, ALT and ATT to represent altitude and attitude) as well as to confusability between pairs of symbols and between pairs of lines on a display such as the navigation display (electronic map).

This module uses the input from the meaningfulness module to "autopopulate" the terms and compute confusability between all possible pairs of terms. Similarly, the ADAT user inputs the symbols used in different FMS displays using a specific tool for creating these symbols. This allows the designer to select from a set of eight shapes and eight colors. As with alphanumeric confusability, symbol confusability is evaluated on a "by display" basis, and pairwise comparisons are made across all symbols within a given display. Symbol

confusability is assessed using a matrix, which assigns penalties for geometric similarity (e.g., triangle and chevron, circle and hexagon) and color similarity (e.g., red and magenta, white and yellow).

2.4.5. Complexity

This module directly addresses FMS functionality, complexity, and the well-documented problems that leave pilots wondering, "What is it doing and why is it doing that?" Somewhat paradoxically, the complexity module is the most complex in terms of its required user inputs and calculations. Input consist of four general elements, based in part on psychological research on cognitive complexity (e.g., Boag, Neal, Loft, & Halford, 2006; Walker, Stanton, Salmon, Jenkins, & Rafferty, 2010) and analyses of the complexity of the FMS itself (Degani, 2004; Hutchins, 1996; Javaux, 2002; Pritchett, 2009). First, the sheer number of modes in a proposed design is an obvious contributor to complexity (Riley, 2001). A large number of modes are not necessarily an overall "bad" design feature, if it offers a more powerful system. However, it does generally increase complexity of the system.

Second, various forms of *interactions*, or *couplings*, between modes contribute to complexity, the facet that Halford, Baker, McCredden, and Bain (2005) refer to as *relational complexity*. This interaction can take on several forms. Activation of one mode may prohibit pilot selection of another mode; it may automatically trigger a transition to a different mode or the invocation of one mode may override previously specified constraints or settings.

Third, complexity is, in part, based on the extent to which the pilot directly controls the mode transitions or setting targets that trigger a transition. Actions and targets that the pilot specifically sets are considered more understandable and less complex than actions and targets selected by automation. The psychological phenomena underlying this penalty is known as the generation effect (Slamecka & Graf, 1978), an observation that people will remember changes in state better if they have actively implemented those changes than if they more passively witnessed another agent making the same change. This principle underlies the success of active learning (e.g., Keith & Frese, 2008).

Fourth, to a considerable extent, the penalties associated with relational complexity can be offset by offering immediate, intuitive graphical *feedback*, regarding what modes are guiding the aircraft and their im-

plications for the future flight path. Correspondingly, absent such feedback, further penalties are imposed. The greatest need for such feedback lies in the vertical modes of flight (i.e., changing altitude) and is present in the vertical situation display (VSD; Hutchins, 1996; Jacobsen, Chen, & Widemann, 2000; Pritchett, 2009; SAE International 2001; Vakil et al., 1995). There are at least two reasons why such a display will enhance situation understanding and reduce the apparent complexity of mode understanding. First, both thrust and pitch can be involved, in different proportions, in changing altitude, depending on the goals of a mode (e.g., economy, speed, comfort), and state (e.g., headwinds, energy state). Second, the three variables of altitude, airspeed, and ground speed are often coupled in complex ways that have major implications for reaching future targets along the flight path. The VSD can make such predictions explicit.

Although the complexity module is not based on a single model of cognitive complexity or intuitive displays, the contribution of these elements to human understanding of automation systems is well documented by research in human–automation interaction. For this module, the designer focuses on each proposed mode individually, designating its purpose, the axes of flight that are controlled, how the mode is invoked, how its target values are set, and a large amount of information regarding the availability and type (e.g., graphic versus text) of feedback. Within this module, the various penalties for each mode are identified, weighted, and summed.

2.4.6. Procedures

This module estimates the "ease of programming," or of "instructing," of the automation to alter its intentions and actions. The module is loosely grounded in the Goals, Operators, Methods, and Selection rules (GOMS) theory of procedural actions in human–computer interaction (Card, Moran, & Newell, 1983; Polson, Irving, & Irving, 1995). A high complexity of programming is clearly an invitation to human error, particularly when feedback from the programming action is not readily visible to the pilot or is infrequently used (Fennell et al., 2006). In addition, the module penalties are also based on fundamental psychological principles of stimulus–response (or display–control) compatibility (Proctor & Van Zandt, 1994; Wickens & Hollands, 2000), as well as those of movement time and accuracy (Fitts, 1954).

TABLE 1. A Summary of the Key Design Issues Identified by Each of the ADAT Modules

Module	Design Factors Evaluated
Display layout	Is the most important / frequent used information centrally located and readily accessible?
	Are displays that need to be used sequentially located near one another?
Changes	Are changes presented in a salient manner?
	Are the most important changes centrally located?
	What is the pilot's anticipated workload at the time a change is expected to occur?
	How frequently is the change expected to occur in flight?
Meaningfulness	Do terms follow human factors principles for acronym and abbreviation formation?
	Are terms used to represent unique concepts?
	Are the terms likely to be understood by pilots, based on computational analyses of aviation terms (using latent semantic analysis)?
Confusability	How similar (in terms of features) are terms and symbols?
	How likely are terms and symbols to be confused with other terms and symbols?
Complexity	How many modes are included in the FMS?
	What axes of flight do they control?
	How are mode transitions initiated?
	What feedback is provided regarding current mode and future mode status?
Procedures	How many steps are required to implement procedures?
	Where are these actions taken?
	Is prompting provided to guide the pilot?
	Is feedback provided? If so, where is it located with respect to the control action?
	Is the sequence of actions consistent with air traffic control clearances?
	Does the action have consequences for other modes?

For this module, the designer specifies the actions required for setting modes and targets, using the action regions identified in the cockpit layout module (Figure 3). On the basis of the sequence, location, prompting, feedback, and complexity of different actions, a number of separate penalties are computed for each action sequence. As in the complexity module, these penalties are tallied, weighted, and summed.

2.4.7. Comparison of Modules

Table 1 shows the six modules included in ADAT and provides a short summary of the key design issues the module identifies.

2.5. Additional Properties of Modules

Although the six modules were designed to be relatively independent from one another, as outlined above, there is some interdependence among them. When the designer enters data for certain modules, this data will autopopulate or automatically carry over to other module data entry fields. For example, a designer specifies displays and their location in the layout module. In the changes module, the designer must identify the display in which each change occurs. The displays selected in the layout module are provided as pull-down menu options in the changes module. Similarly, each pilot-initiated mode selection (identified in the complexity module) is associated with a procedure. All pilot-initiated mode selections appear in the procedures module as procedures that need to be defined. To the extent possible, these cross-links are implemented automatically by ADAT, mitigating the need for the designer to enter corresponding data repeatedly across different modules.

After the designer has filled the appropriate data-entry screens for each module, the designer clicks the "evaluate" button, and within seconds, the computational penalties and score of the design are displayed. At that point, it is possible to return to the data-entry screens, modify the design, and recompute the evaluation. Individual modules provide an overall score, as well as a set of subscores. The subscores relate to individual components being evaluated (e.g., each individual control mode in the complexity module, each individual procedure, or each pair of potentially confusable symbols or terms).

2.6. Research-Based Guidance

The previous description has highlighted the documented *problems* (or penalties) associated with different design modules. Critically, however, ADAT also provides users with both guidelines on how to mitigate problems and remedies for improving a low score, as well as documented research to provide a background regarding the source of the penalties or efficacy of the guidelines.

The ADAT library of research-based guidance includes more than 60 references, presented as abstracts and summaries (summaries are more detailed than abstracts but nonetheless are shortened versions of the original article). The references are generally empirical research of human performance on the flight deck, and many specifically investigate FMS design issues. Three broader documents include FAA / European Aviation Safety Agency specifications, the FAA Human Factors Design Standard (Ahlstrom & Longo, 2003), and recommendations from the Society of Automotive Engineers (2007) for flight deck design. Actual guidelines are taken from the first two sources, which are publicly available; the SAE guidelines are summarized.

There are three primary ways a designer would access these references. The first is as a reference library, when beginning a new design. The FMS designer can type in a search term of interest, and identify all relevant articles that address the design issue. The second way of accessing relevant literature is through the ADAT Help system, provided within each module. This gives the designer a list of all literature relevant to that particular design module. Third, the most typical way of accessing the literature will be through feedback and specific literature references presented when module evaluations are performed.

In developing ADAT, we identified the possible patterns of results that each module could provide. We then associated these results with particular "bullet-point guidance"—concise recommendations for improving the design. Further, we identified relevant literature for each of the results and included hyperlinks to this literature on the results page. The literature either provides more information (generally research but also aviation human factors or industry standards) on the particular problem identified or how the designer can resolve the problem.

2.7. Attention and Noticing Models: SEEV and N-SEEV

In addition to the six modules, described above, ADAT evaluates the proposed FMS design by simulating the expected pilot scanning behaviors and predicting the likelihood that pilots will notice specific changes or events on the flight deck. These predictions are made using the SEEV (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003; Wickens & McCarley, 2008) and N-SEEV attention models (McCarley et al., 2009; Steelman-Allen et al., 2009). Accordingly, SEEV predicts that the visual scan will be attracted to areas that are salient (S), will be inhibited to the extent that effort (E) is required to scan longer distances, will travel to locations where information is expected (E) because of a high bandwidth or event rate at the source, and will travel to locations of higher value (V), or relevance, to the task in question.

The output from SEEV is the predicted visual scan, and the percent of time a pilot is expected to look at each display in the visual field. N-SEEV predicts the pilot's ability to *notice* changes in the visual world. SEEV factors affect pilot steady-state scanning behavior, and once a change occurs, the pilot's speed of noticing will depend on the extent to which workload is lower; the event to be noticed is salient, expected, valuable, and is located relatively closer (less eccentric) to the momentary direction of scan when the event occurs.

If too much time has elapsed before the location of the change event is fixated—typically around 8–10 s, although this can vary (Wickens, Hooey et al., 2009)—the well-known problem of change blindness is likely to occur (Rensink, 2002). N-SEEV thus provides data on the probability that the pilot will notice a specific change (e.g., flight mode annunciator onset) and predicts the time required to notice that change.

ADAT uses aspects of SEEV and N-SEEV to evaluate the proposed FMS design as part of the layout and changes modules. However, the output provided by the attention models, shown in Figure 4, is qualitatively different from the two module scores, because the attention models show how the proposed design supports (or negatively impacts) pilot noticing behavior and performance using a dynamic real-time scan simulation across the flight deck. In addition, the attention models produce graphs that show the percent of time the pilot is expected to focus on each display
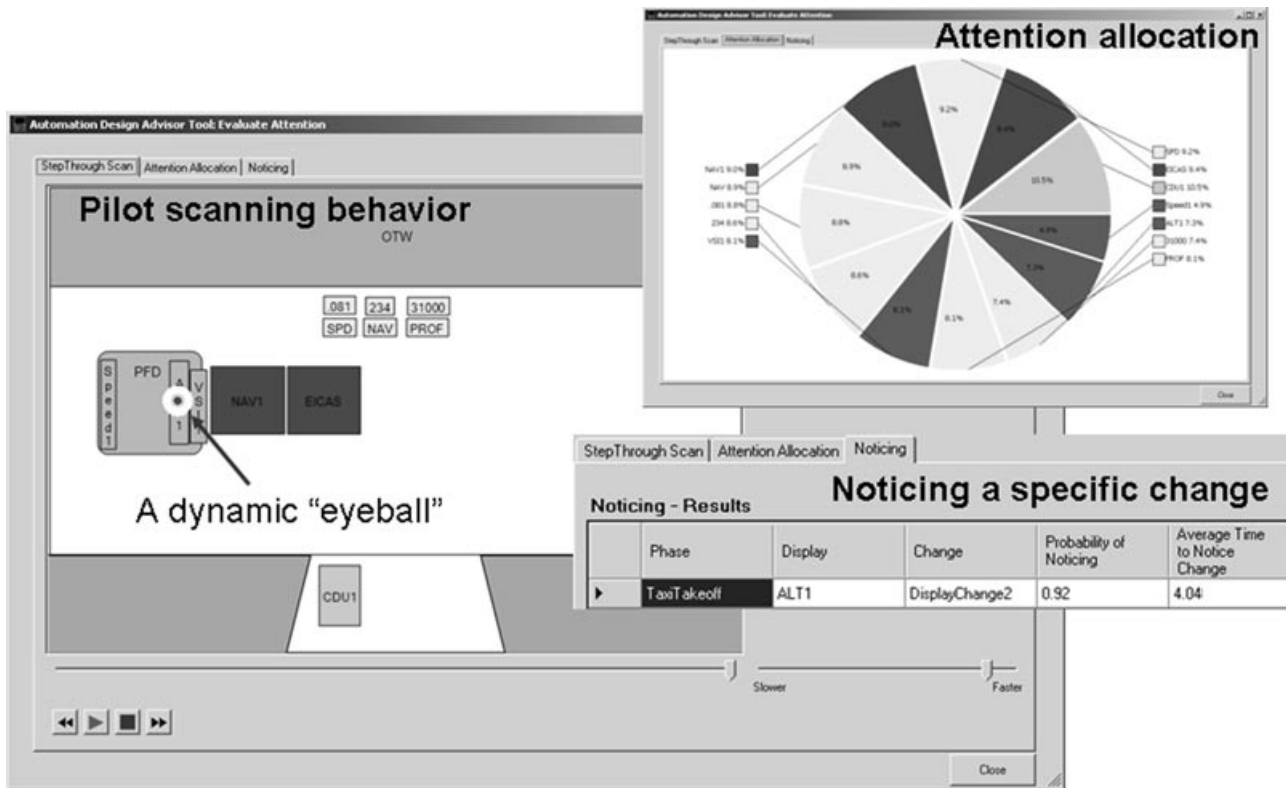
Figure 4 ADAT attention model outputs—the predicted scan pattern, attention allocation, noticing probability, and time required to notice a change.

and can reveal display areas that may be neglected in the typical pilot's scan pattern. They also generate predictions regarding the probability and time to notice a specific change of interest.

3. ADAT USES

### 3.1. Flight Deck Design and Comparing Design Options

We expect that ADAT can be used by FMS designers and avionics manufacturers to identify and remediate potential design deficiencies and to compare potential FMS designs. ADAT was specifically designed to support these comparisons. The "compare" utility allows designers to identify up to three previously specified FMS designs and to compare them on a feature-by-feature basis. Figure 5 shows two screen shots of this utility in which two FMS designs are compared. The figure on the left shows the high-level comparison of two designs. The features evaluated by the modules are identified in the first column, and each of the two designs is identified in the remaining columns. A "stoplight status" indicator (i.e., red indicates design prob-

lems, yellow indicates potential concerns, and green indicates a strong design as predicted by the ADAT scores) provides a high-level overview of the FMS design score for each of the module evaluation areas for each design. This table gives designers an at-a-glance comparison of the two designs.

The figure on the right shows a more detailed window. When the designer clicks on a tab, he or she can access the detailed scores for each of the tab issues. Detailed results for each design are presented on subtabs within the module tab. The designer can easily see the specific results for each design and identify which specific FMS design features best support pilot performance. This will help designers and managers decide how to prioritize efforts on upgrades and how to select among potential design alternatives.

3.2. Input for FAA Certification

We expect that ADAT evaluations can be helpful to the FAA in evaluating potential new FMS designs for flight certification (Abbott et al., 1996; European Aviation Safety Agency, 2007). Guidance for addressing identified design problems is directly associated with
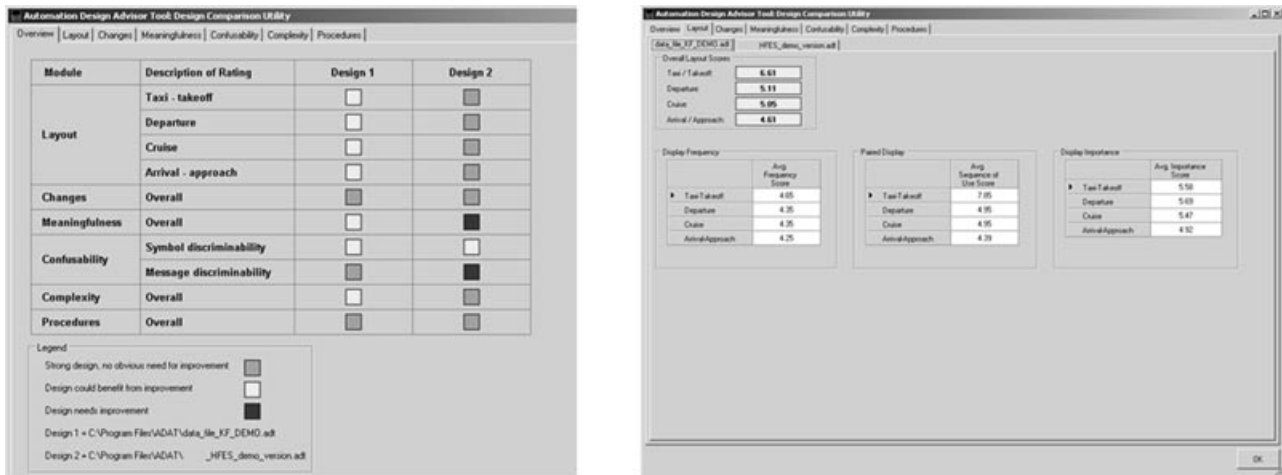
Figure 5 The ADAT design comparison utility, showing a comparison of two FMS designs. The screen on the left shows the overall design comparison, and the screen on the right shows the tabs that provide a detailed, per module comparison.

specific elements in both the EASA document and the FAA's Human Factors Design Standard (Ahlstrom & Longo, 2003). By providing a tool to evaluate the design in terms of its ability to support pilot performance, we enable FAA evaluators to identify potential concerns with new designs. Further, because ADAT supports easy comparisons across designs, the tool allows FAA personnel to compare a NextGen design against a current-day design. In our reviews with both avionics designers and discussions with FAA personnel, we received feedback that the tool appears particularly useful by offering thorough, systematic methods of evaluating and comparing design concepts.

4. USABILITY AND VALIDATION STUDIES

4.1. Approach

Our approach to usability and validation testing included three main tasks: 1) conducting heuristic-based usability evaluations, 2) performing module-by-module validations to assess a particular aspect of design, and 3) carrying out overall tool validations. These tasks were performed multiple times throughout the design process. This approach allowed us to update and improve the tool during the design process and enabled us to conduct subsequent evaluations on a progressively more refined product. The methods and results of each of these validation efforts are described individually.

4.2. Usability Evaluations

Heuristic-based usability evaluations were performed by a team member dedicated to that task. The evaluator's role was to step through the tool and identify inconsistencies and inefficiencies in the way tasks were implemented. The evaluator also reviewed the help system to see how well it supported users in performing an FMS evaluation.

Usability heuristic evaluations helped ensure consistent functionality across the modules. These evaluations identified where we could reduce the number of steps ADAT users needed to perform to input data or complete tasks. We identified multiple areas in the tool where data input for one module could be carried over to another module, as described previously. The tool was structured so that users complete the data entry by following the tabs left to right (see Figure 3, top left), allowing ADAT to autopopulate later modules with input from the earlier modules. These automated data-entry improvements significantly reduced the time required to enter an FMS design. We noted a frequent use of distracting pop-up windows, so we replaced them with data-entry tabs. One significant improvement of the tool was to restructure the "help" system to match the way users are likely to work with the tool. Users are able to find quick answers to specific questions (e.g., by buttons labeled with question marks to indicate "more information available") and to obtain specific, step-by-step guidance on how to work with the module.

4.3. Individual Module Validations

Throughout the ADAT development process, we conducted numerous small-scale validation studies. Initially, we met with SMEs to identify and to characterize problems in FMS design. As we developed individual modules, we met again with these personnel to discuss our approach and to obtain their input and feedback. These discussions provided valuable insights into concerns that we had either neglected or minimized and helped us to properly account for the factors they considered most critical.

Once a module was developed and implemented to the point where it could assess and provide scores of FMS design issues, we met again with SMEs to show what had been developed, discuss the design issues being evaluated, and indicate how the module assessed performance. The number of meetings and total SMEs interviewed varied depending on the importance of the specific module (e.g., the complexity module, which addresses perhaps the most crucial aspect of FMS usability, included the greatest number of reviews and involved the greatest number of aviation SMEs). The primary focus of these validations was on the relevance and weightings of the calculations to assess FMS design quality (e.g., did the factors we identified and the required designer inputs accurately capture those design issues related to complexity?). During several evaluations, we also requested an aviation SME to work with the tool and identify questions or usability concerns, many of which were addressed by simple design changes. Some of our SMEs were also instructors who had taught student features of the commercial FMS, and hence, they were particularly sensitive to problems that less-skilled pilots would have in using flight deck automation. In these individual module evaluations, we identified potential concerns with the algorithms and modified them to provide more reasonable, realistic evaluations of FMS design quality.

During module validation discussions, we also found that aspects of the ADAT tool design (e.g., questions, terms) were not readily apparent to the aviation SMEs. To eliminate this potential for confusion, either we changed the wording of these terms or data entry questions or we provided more detailed guidance on the relevant ADAT module page or in the help system.

4.4. Overall Validation Analyses

Once a complete ADAT prototype was developed, we began testing by evaluating actual and potential future FMS designs. For all of these validation studies, we provided an overview of ADAT, explaining how it works, what inputs are required, and what the scores indicate. Prior to starting the review, the evaluators (pilots, FMS designers, and researchers) gathered data describing FMS designs, including training and operating manuals, design plans, or publications summarizing the design. The reviewers then stepped through the ADAT tool, module by module, entering the data describing the FMS design.

In our ADAT-based evaluations, we compared the quality of design between a conventional FMS, existing on three different aircraft: the MD11, a Boeing 777, and a Gulfstream 550, with a true "revolutionary" design of the Flight Deck of the Future (FDF: Mumaw, Boorman, & Prada, 2006). We were also interested in comparing across the design of the more conventional automated aircraft (Boeing 777) with the newer, but not revolutionary, design of the Gulfstream 550. Table 2 provides details of the ADAT-based evaluations of different flight deck automation systems and depicts the scores for the different modules.

Before describing the details of these evaluation scores and the score comparisons that follow, it is important to distinguish the different kinds of contrasts that are made, as well as particular features within each

TABLE 2.   ADAT Scores for Each Evaluation

	MD11	FDF UM	FDF Alion	Boeing 777	GS 550
Layout Changes	8.1	8.4	9.0	6.1	7.5
Meaningfulness	4.0	7.1	5.4	5.9	3.6
Confusability	9.4	9.7		7.6	7.6
Complexity				7.5	6.4
Procedures	4.4	6.3	8.1	5.4	5.7

evaluation. In all contrasts but one (the MD11 and the FDF UM), the evaluators were different people. The MD11 and FDF UM were evaluated by one rater, a commercial pilot and graduate student in human factors. The FDF Alion was performed by a human factors professional. The Boeing 777 evaluation was done by a professional pilot and flight instructor for advanced automated systems. The GS550 was conducted by a human factors professional avionics designer. Finally, not all evaluators evaluated all modules. In particular the FDF design was only intended by its developers to provide part of the functionality of an FMS and did not include the display regions (e.g., navigation display, or VSD) that would normally present a good deal of flight deck information and be part of a full ADAT evaluation. Considering these caveats, six comparative evaluations reveal the following results (details are reported in Sarter, Wickens, & Sebok, 2011).

4.4.1. Revolutionary versus Conventional Systems

1. FDF UM and MD11. In contrasting the FDF with equivalent aspects of functionality on the MD11 FMS (e.g., that functionality required to supervise and program the automation), the FDF produced a superior score on all modules, with large gains related to meaningfulness of terms and procedures. The "revolutionary" FMS, designed by human factors professionals with human factors issues very much in mind, was rated by ADAT as a superior design.

2. FDF Alion and MD11. In the same contrast as (1), except with a different rater, the ADAT evaluation revealed higher scores for the FDF than the MD11. While there were differences in the FDF evaluations between the two raters (discussed below), the trend of higher scores for the FDF design is consistent.

3. FDF UM and Boeing 777. In this contrast, the FDF was compared with a "baseline" aircraft. In this case, the baseline was the Boeing 777, a somewhat newer design than the MD11. In this comparison, like the ones described previously, the ADAT scores were higher for the FDF. layout, meaningfulness, confusability, and procedures were all found to be improved in the FDF compared with the Boeing 777.

4. FDF Alion and Boeing 777. This evaluation also revealed an advantage for the FDF for the Layout and Complexity modules, but the not for meaningfulness of terms. This difference might reflect greater attention to the meaningfulness of terms in the more recent Boeing 777 design, compared to the MD11.

4.4.2. Conventional Older versus Newer Systems

5. Boeing 777 and the GS550. In this contrast, the design of the Boeing 777 FMS was compared with that of the GS550. The GS550 is a more recent design, and, being a corporate jet, it is freed from some of the developmental constraints required of commercial airlines designs. Here the comparative evaluation reveals few substantial differences in the module scores. We note here that the ratings for the confusability module are lower for both the Boeing 777 and the GS550 compared with the MD11 and FDF (UM). This is because in the Boeing 777 and GS550 evaluations confusability was evaluated for both terms and symbols. In the MD11 and FDF (UM) analyses, only term confusability was evaluated. In all designs, terms were highly discriminable, whereas symbols were potentially confusable.

4.4.3. Same System, Different Raters

6. FDF UM and FDF Alion. In contrasts 1–4, described above, the differences in scores could have resulted, to some extent, from both actual design differences between systems and from what might be described as "inter-rater reliability." That is, when different raters are involved, they might make different assumptions about how data should be input into ADAT. The best evaluation of this measure of inter-rater reliability is provided by contrasting the FDF (UM) and FDF (Alion) columns of Table 2, where the same system (FDF) was evaluated by two different raters. Here we note substantial disagreement on the three modules rated by both raters. Detailed scrutiny of the sources of these differences, reported in Sarter et al. (2011) reveals a number of cases of evaluators making different assumptions

TABLE 3. Comparison of the FDF and a Conventional FMS

	Training Accuracy Session 2	Transfer Session 1	Transfer Session 2
FDF	69%	86%	68%
Generic FMS	57%	65%	53%

Note: Experimental data as reported in the original work by Mumaw et al. (2006).

regarding the same system. For example, sometimes the same abbreviation in the meaningfulness module was assigned to a different coding category by the two different raters (e.g., truncation or deletion). Evaluators also differed in their willingness to extrapolate from documentation to the actual functionality. In this particular case, the FDF is not an implemented system but is a design concept described in research papers, making it particularly vulnerable to differences in interpretation. These differences led to elaboration of the ADAT help system to improve consistency across evaluators.

4.4.4. Validation against Empirical Performance

7. FDF versus "conventional" FMS. Ultimately, the best validation of ADAT will be obtained if the model predicts differences in actual pilot performance using different versions of the FMS. Only one such data set was available to us, a comparison by Mumaw et al. (2006) of programming either the FDF or a conventional, current-day FMS. Ten different FMS programming tasks were carried out by pilots on computer simulations of each of the two different flight deck automation designs. An evaluation was then carried out on several measures across two different sessions, the results shown in Table 3. Note that the difference between performance with the FDF compared with a generic FMS was significant at the $p < .05$ level only for Transfer Session 1. For the other two conditions (Training Accuracy Session 2 and Transfer Session 2), the differences were not significant.

The measures of training and transfer were essentially measures of programming actions and hence correspond to the procedures module predictions. If the Boeing 777 could be considered an example of the "generic FMS," then these performance scores could be compared with the two procedures module ADAT scores. From Table 2, the two FDF procedures scores were 6.3 (UM) and 8.1 (Alion). The Boeing 777 procedures score was 5.4. Averaging the two FDF scores gives an overall rating of 7.2. Comparing this to the Boeing 777 score (7.2 compared with 5.4), we find the FDF scores 33% higher. Interestingly, if the three scores that represent empirical performance with procedures in Table 3 are averaged (Training Accuracy, Transfer Sessions 1 and 2), the actual performance during evaluation was 27% higher for the FDF relative to the reference system.

4.4.5. Validation Summary

The validation studies indicate that ADAT evaluations distinguish among different types of FMS designs and that the differences are generally what would be expected. A novel, revolutionary FMS, the FDF, with integrated, predictive information, shows higher layout scores compared with traditional FMS designs. Similarly, comparing ADAT results with an empirical study suggests that ADAT evaluations accurately assess which system supports better performance, and even approximately predict the degree of improvement. The tool is sensitive to differences in user assumptions, providing different ratings of a single system depending on user inputs. Further research is needed to investigate, validate, and improve ADAT. This should include detailed analyses of FMS designs and corresponding ADAT evaluations to be compared with empirical studies of pilot performance using these systems.

5. CONCLUSION

The transition to NextGen operations will offer numerous challenges to aviation designers. As automation assumes an even more prominent role on the flight deck, designers are tasked with ensuring that automation adequately supports the pilot. ADAT can provide a useful tool to help designers proactively identify and resolve potential design deficiencies that impact human–automation interaction.

ACKNOWLEDGMENTS

guidance on this research. In addition, subject-matter experts Karl Fennell, Rick Shay, Ron Small, Clay Hubbs, Emil Lassen, Mark Sims, John Hillier, Dan Lewis, and Nick Johnson all provided valuable inputs. Dr. Randy Mumaw and Dr. Vic Riley kindly clarified Flight Deck of the Future concepts for the team. Dr. Marilyn Blackmon and Dr. Peter Polson graciously provided access to their LSA capability. This work was performed for NASA Ames and the Integrated Intelligent Flight Deck program, under contract NNX07AT79A.

References

Abbott, K., Slotte, S., Stimson, D., Bollin, E., Hecht, S., & Imrich, T. (1996). The interfaces between flightcrews and modern flight deck systems (Federal Aviation Administration Human Factors Team Report). Washington, DC: Federal Aviation Administration.

Ahlstrom, V., & Longo, K. (2003). Human Factors Design Standard (HF-STD-001). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representation in natural tasks. Journal of Cognitive Neuroscience, 7(1), 66–86.

Blackmon, M. H., Mandalia, D. R., Polson, P. G., & Kitajima, M. (2007). Automating usability evaluation: Cognitive walkthrough for the web puts LSA to work on real world HCI design problems. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), Handbook of latent semantic analysis (pp. 345–375). Mahwah, NJ: Lawrence Erlbaum.

Boag, C., Neal, A., Loft, S., & Halford, G. S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. Ergonomics, 49(14), 1508–1526.

Boorman, D. J., & Mumaw, R. J. (2004). A new autoflight/FMS interface: Guiding design principles. Proceedings of the International Conference on Human-Computer Interaction in Aeronautics, Toulouse, France.

Card, S., Moran, T., & Newell, A. (1983). The psychology of human-computer interaction. Hillsdale, NJ: Lawrence Erlbaum.

Degani, A. (2004). Taming Hal: Designing interfaces beyond 2001. New York: Palgrave Macmillan.

Driskell, J. E., & Salas, E. (1996). Stress and human performance. Mahwah, NJ: Lawrence Erlbaum.

European Aviation Safety Agency. (2007). Certification specifications for large aeroplanes. CS-25. Book 2: Acceptable Means of Compliance.

Fennell, K., Sherry, L., Roberts, R. J. Jr., & Feary, M. (2006). Difficult access: The impact of recall steps on flight management system errors. International Journal of Aviation Psychology, 16(2), 175–196.

Fiorino, F. (2006). TCAS a key factor in Amazon midair. Aviation Week and Space Technology. October 8, 2006.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 47, 381–391.

Funk, K., Lyall, B., Wilson, J., Vint, R., Niemczyk, M., Suroteguh, C., & Owen, G. (1999). Flight deck automation issues. International Journal of Aviation Psychology, 9(2), 109–123.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? Psychological Science, 16(1), 70–76.

Hutchins, E. (1996). The integrated mode management interface. Final report for grant #NCC 2-591 for NASA. Ames, IA: Aviation Safety / Automation Program.

Jacobsen, A., Chen, S., & Widemann, J. (2000). Vertical situation awareness displays. Proceedings Royal Aeronautical Society Conference on Situation Awareness, London, UK.

Javaux, D. (2002). A method for predicting errors when interacting with finite state systems. How implicit learning shapes the user's knowledge of a system. Reliability Engineering and System Safety.75(2), 147–165.

John, B. E., Blackmon, M. H., Polson, P. G., Fennell, K., & Teo, L. H. (2009). Rapid theory prototyping: An example of an aviation task. Proceedings of the Human Factors and Ergonomics Society the 53rd Annual Meeting, October 19–23, 2009. Santa Monica, CA: Human Factors and Ergonomics Society.

Joint Planning and Development Office (JPDO). (2010, April 10). Concept of operations for the next generation air transportation system. Version 3.1. Washington, DC: Federal Aviation Administration.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. Journal of Applied Psychology. 93(1), 59–69.

Landauer, T. K. (1995). The trouble with computers: Usefulness, usability, and productivity. Cambridge, MA: MIT Press.

McCarley, J., Wickens, C., Sebok, A., Steelman-Allen, K., Bzostek, J., & Koenecke, C. (2009). Control of attention: Modeling the effects of stimulus characteristics, task demands, and individual differences. NASA Final Report for NASA ROA 2007 Integrated Intelligent Flight Deck Technologies, Contract No. NNX07AV97A.

Mumaw, R., Boorman, D. J., & Prada, R. L. (2006). Experimental evaluation of a new autoflight interface. Proceedings HCI-Aero 2006, International Conference on Human Computer Interaction, September 20–22, 2006, Seattle, Washington.

NextGen Integration and Implementation Office. (2010). NextGen implementation plan. Washington, DC: Federal Aviation Administration.

Nikolic, M., Orr, J., & Sarter, N. (2004). Why pilots miss the green box: How display context undermines attention capture. International Journal of Aviation Psychology, 14(1), 39–52.

Polson, P. G., Irving, S., & Irving, J. E. (1995). Applications of formal models of human-computer interaction to training and the use of the control and display unit [Final report]. Washington, DC: Federal Aviation Administration.

Prada, L. R., Mumaw, R., Boehm-Davis, D., & Boorman, D. (2006). Testing Boeing's Flight Deck of the Future: An experimental comparison of prototype and current automated flight panels. Proceedings of the 50th Human Factors and Ergonomics Society Annual Meeting, San Francisco, CA, October 2006.

Pritchett, A. (2009). Aviation automation: General perspectives and specific guidance for the design of modes and alerts. In F. Durso (Ed.), Reviews of human factors and ergonomics (Vol. 5, pp. 82–113). Santa Monica, CA: Human Factors and Ergonomics Society.

Proctor, R. W., & Van Zandt, T. (1994). Human factors in simple and complex systems. Needham Heights, MA: Allyn & Bacon.

Rensink, R. A. (2002). Change detection. Annual Review of Psychology. 53(4), 245–277.

Riley, V. (2001). A new language for pilot interfaces. Ergonomics in Design. 9(2), 21–26.

SAE International. (2001). Human Interface criteria for vertical situation awareness displays. ARP 5430. SAE.

Sarter, N., Wickens, C. D., & Sebok, A. (2011). Proactive system design and evaluation: supporting pilot-automation interaction through empirical and modeling analyses. NASA Final Report (Contract NNX07AT79A).

Sarter, N., & Woods, D. D. (2000). Team play with a powerful and independent agent: A full-mission simulation study. Human Factors 42(3), 390–402.

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on highly automated flight decks: An empirical study combining behavioral and eye tracking data. Human Factors, 49, 347–357.

Sarter, N. B., & Woods, D. D. (1992). Pilot interaction with cockpit automation: Operational experiences with the flight management system (FMS). International Journal of Aviation Psychology, 2(4), 303–321.

Sarter, N. B., & Woods, D. D. (1994). Pilot interaction with cockpit automation: II. An experimental study of pilots' models and awareness of the flight management system. International Journal of Aviation Psychology, 4, 1–28.

Sherry L., & Polson, P. G. (1999). Shared models of FMS vertical guidance. International Journal of Aviation Psychology, 9, 139–154.

Sherry, L., Feary, M., Polson, P., & Fennell, K. (2003). Drinking from the fire hose: Why the flight management system can be hard to train and difficult to Use. NASA Technical Report: NASA/TM-2004–212274.

Sherry, L., Fennell, K., Feary, M., & Polson, P. (2006). Analysis of ease-of-use and ease-of-learning of a modern flight management system. AIAA Journal of Aerospace Computing, Information, and Communication, 3, 177–186.

Sherry, L., Fennell, K., Feary, M., & Polson, P. (2006b). Analysis of flight management system messages. AIAA Journal of Aircraft, 43, 1372–1376.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: Human Learning and Memory, 4(6), 592–604.

Society of Automotive Engineers. (2007). Aerospace recommended practice: Flight Deck Alerting System (FAS). SAE ARP4102/4.

Steelman-Allen, K. S., McCarley, J. S., Wickens, C., Sebok, A., & Bzostek, J. (2009, October). N-SEEV: A computational model of attention and noticing. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 53, 774–778.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327–352.

Vakil, S., Hansman, R. J., & Midkiff, A. (1995). Impact of vertical situation information on vertical mode awareness in advanced autoflight systems. 14th AIAA/IEEE Digital Avionics Systems Conference, November 1995.

Walker, G. H., Stanton, N. A., Salmon, P. M., Jenkins, D. P., & Rafferty, L. (2010). Translating concepts of complexity to the field of ergonomics. Ergonomics, 53(10), 1175–1186.

Wickens, C., Hooey, B., Gore, B., Sebok, A., & Koenecke, C. (2009). Identifying black swans in NextGen: Predicting human performance in off-nominal conditions. Human Factors, 51, 638–651.

Wickens, C. D. (2003). Aviation displays. In P. Tsang & M. Vidulich (Eds.), Principles and practice of aviation psychology (pp. 147–199). Mahwah, NJ: Lawrence Erlbaum.

Wickens, C. D. (2010). The psychology of aviation surprise: An 8-year update regarding the noticing of black swans [Keynote address]. International Symposium on Aviation Psychology, Wright State University, Dayton, Ohio.

Wickens, C. D., & Alexander, A. L. (2009). Attentional tunneling and task management in synthetic vision

displays. International Journal of Aviation Psychology, 19, 182–189.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. Human Factors, 37(3), 473–494.

Wickens, C. D., & Hollands, J. (2000). Engineering psychology and human performance (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., & McCarley, J. M. (2008). Applied attention theory. Boca Raton, FL: CRC Press / Taylor & Francis.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W., & Talleur, D. A. (2003). Attentional models of multi-task pilot performance using advanced display technology. Human Factors, 45(3), 360–380.

Wickens, C. D., McCarley, J. M., Steelman-Allen, K., Sebok, A., Bzostek, J., & Sarter, N. (2009). NT-SEEV: A model of attention capture and noticing on the flight deck. Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society, October 19–23, 2009, San Antonio, Texas.

Wickens, C. D., Vincow, M. A., Schopper, A. W., & Lincoln, J. E. (1997). Computational models of human performance in the design and layout of controls and displays (CSERIAC SOAR Report 97–22). Wright Patterson Air Force Base, OH: Crew System Ergonomics Information Analysis Center.

Wiener, E. L. (1989). Cockpit automation. In E. L. Wiener & D. C. Nagel (Eds.), Human factors in aviation (pp. 433–461). New York: Academic Press.