NASA/TM–2014-218661

Interval Management Display Design Study

Brian T. Baxley, Timothy M. Beyer, Stuart D. Cooke, and Karlus A. Grant
Langley Research Center, Hampton, Virginia

December 2014

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM–2014-218661

Interval Management Display Design Study

Brian T. Baxley, Timothy M. Beyer, Stuart D. Cooke, and Karlus A. Grant
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

December 2014

Acknowledgments

This NASA Technical Memorandum documents the methodology and results of a human-in-the-loop study about new cockpit displays for pilots to conduct Interval Management (IM) operations. The authors gratefully thank Sara Wilson for providing critical guidance for the study design and data analysis, as well as reviewing this document. The authors also thank Stella Harrison for the many hours she invested to enable this study.

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

ASTAR	Airborne Spacing for Terminal Arrival Routes
ATC	air traffic control
ATD-1	Air Traffic Management Technology Demonstration–1
ATOS	Airspace and Traffic Operations Simulation
CGD	configurable graphics display
EAGAR	EcoDemonstrator ASTAR Guided Arrival Approach
EFB	Electronic Flight Bag
FAA	Federal Aviation Administration
FIM	Flight Deck-based Interval Management
FOV	field of view
IM	Interval Management
KMWH	Grant County International Airport
KPHX	Phoenix Sky Harbor International Airport
NASA	National Aeronautics and Space Administration
NextGen	FAA's Next Generation Air Transportation System
SME	subject matter expert
SRD	Systems Requirement Document

1. Introduction

1.1. Background

In 2012, the Federal Aviation Administration (FAA) estimated that U.S. commercial air carriers moved 736.7 million passengers over 822.3 billion revenue-passenger miles (ref. 1). The FAA also forecasts, in that same report, an average annual increase in passenger traffic of 2.2 percent per year for the next 20 years, which approximates to one-and-a-half times the number of today's aircraft operations and passengers by the year 2033. If airspace capacity and throughput remain unchanged, then flight delays will increase, particularly at those airports already operating near or at capacity. Therefore it is critical to create new and improved technologies, communications, and procedures to be used by air traffic controllers and pilots.

National Aeronautics and Space Administration (NASA), the FAA, and the aviation industry are working together to improve the efficiency of the National Airspace System and the cost to operate in it in several ways, one of which is through the creation of the Next Generation Air Transportation System (NextGen). NextGen is intended to provide airspace users with more precise information about traffic, routing, and weather, as well as improve the control mechanisms within the air traffic system. NASA's Air Traffic Management Technology Demonstration-1 (ATD-1) Project is designed to contribute to the goals of NextGen, and accomplishes this by integrating three NASA technologies to enable fuel-efficient arrival operations into high-density airports (ref. 2). The three NASA technologies and procedures combined in the ATD-1 concept are advanced arrival scheduling, controller decision support tools, and aircraft avionics to enable multiple time deconflicted and fuel efficient arrival streams in high-density terminal airspace.

1.2. ATD-1 Project Goal and Concept of Operations

One of the ATD-1 Project's goals is to improve the precision of the spacing between arriving aircraft, thereby increasing capacity at high-density airports and improving aircraft fuel efficiency in the surrounding terminal airspace.

The ATD-1 Concept of Operations begins with the advanced arrival schedule software calculating a conflict-free, time-deconflicted flight plan for all aircraft arriving to that airport (ref. 2). When an aircraft crosses the freeze horizon for that airport (tailored to each airport, ranges from 120 to 250 nautical miles out), the ground scheduling system (the first component of the ATD-1 concept) assigns that aircraft its landing runway and scheduled time of arrival. Controllers then use their decision support tools (the second component of the ATD-1 concept) to assign aircraft an airspeed to meet that scheduled time of arrival. For those aircraft equipped with Interval Management (IM) software (the third component of the ATD-1 concept), the controller has the option to inform the flight crew of the preceding aircraft's call sign, arrival route, and the time interval between the two aircraft.

This study focuses on just the initial pilot response of those aircraft that are IM equipped, and the action they must take to carry out the controller's instruction. The pilot verbally reads back the IM instruction, then enters that information into the onboard IM avionics, which calculates the airspeed needed to meet the schedule (as opposed to the ground system calculating the airspeed for those aircraft not equipped). The algorithm within the onboard IM avionics that calculates the airspeed is called Airborne Spacing for Terminal Arrival Routes (ASTAR).

1.3. Experiment Purpose

Recent experiments conducted at NASA Langley about the ATD-1 operations and the IM procedures and associated displays have created a list of suggested modifications to the displays by the study subjects (typically highly-experienced commercial airline pilots) (refs. 3–9). Additionally, the ATD-1 Project recently released a Systems Requirement Document (SRD) that specified several new capabilities for the IM procedure and the displays. This Project SRD is based on the not yet published draft of the Safety, Performance and Interoperability Requirements Document for Airborne Spacing–Flight Deck Interval Management (ref. 10). Finally, another human-in-the-loop study exploring these expanded IM procedures has been scheduled for 2015 (Interval Management Alternative Clearances, or IMAC), creating an urgency to respond to the previous research results and new requirements. This IM display design study attempted to address the previous research results and suggestions and meet the new requirements, then conduct a rudimentary study using non-commercial airline pilots to determine if the changes were useful. Reference 11 describes the redesign of the IM logic, messages, and displays that were used in this study.

2. Study Methodology

2.1. Objective

The study was a comparison of two IM display interfaces: the current IM system which supports one of the five IM operations (ref. 10), and a revised IM display system which supports four of the five IM operations.

While the current system represents what has been used in NASA's ATD-1 research the past three years, the prototype tool was designed to (1) address research results from those experiments, (2) comply with new requirements and IM clearance types (ref. 10), (3) minimize clearance entry times and errors, (4) provide situation awareness by displaying only necessary information, and (5) evaluate whether the revised IM logic was complete and correct.

2.2. Hardware and Software

Experiment participants used a standard Windows-based laptop, configured with a wired or wireless mouse. The use of a mouse to interface with both of the IM displays in this study is a significant deviation from the normal touch-screen interface used in real world aircraft operations, therefore the relative results between the display types is informative, but the absolute values (time of entry, etc.) are not.

The computers were also loaded with HyperCam2 (internet freeware), which was used to record the monitor's video signal if needed during post-analysis.

To emulate the current IM display system used in ATD-1 research, the EcoDemonstrator ASTAR Guided Arrival Approach (EAGAR) tool was used. This stand-alone module of the Airspace and Traffic Operations Simulation (ATOS) software fully replicates the appearance and action of the current system. Specifically, EAGAR simulates the two displays used in the cockpit for IM operations: the Electronic Flight Bag (EFB) and the configurable graphics display (CGD).

For this study, only the data entry into the EFB was used, and there was no interaction with or questions about the EAGAR CGD. The EFB is to the left in figure 1, and the CGD is to the upper right. The EAGAR tool contained information only for the Grant County International Airport in Moses Lake, Washington (KMWH). Once data entry was complete, there was no connectivity or other software that generated information required to calculate and display the IM speed.

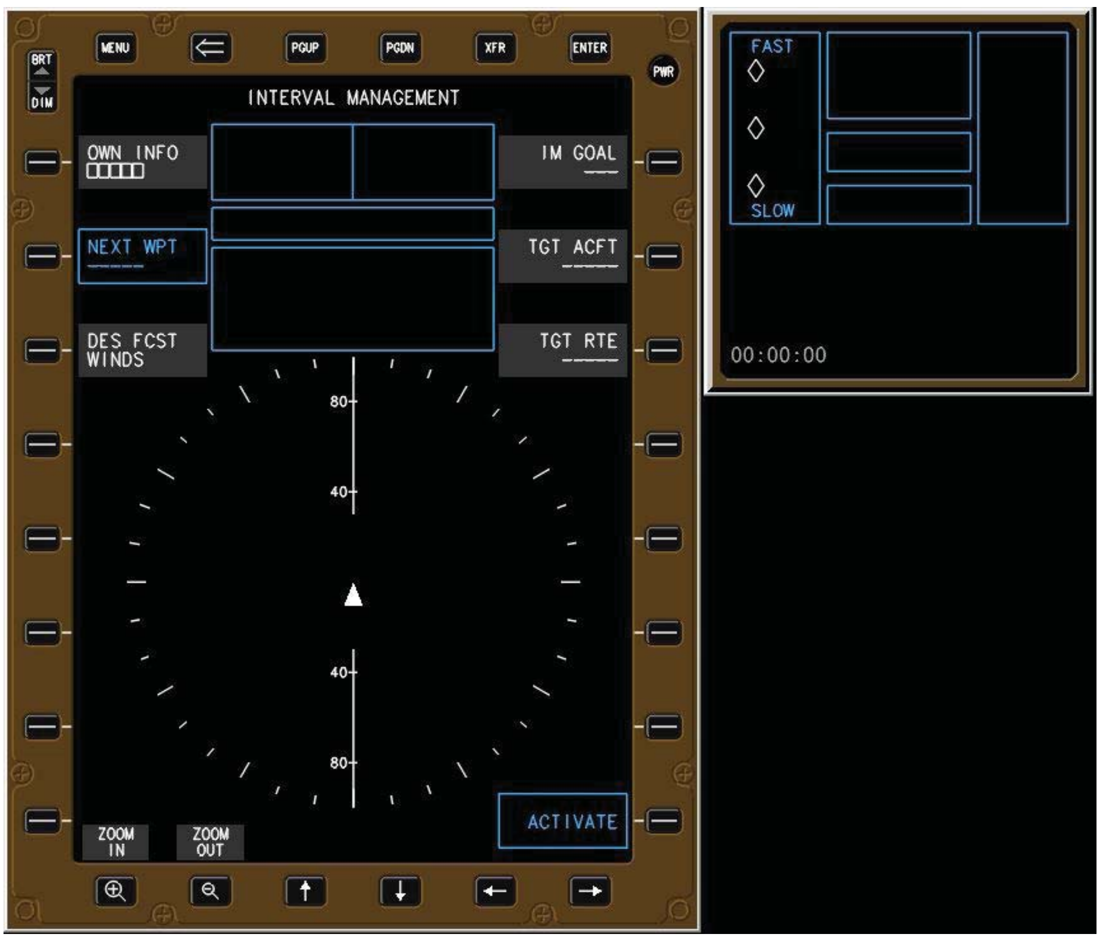

Figure 1. Current IM display as created by the EAGAR tool.

The prototype IM display tool was created by three NASA Langley summer interns (the co-authors of this study), and was done in two phases. First, each individual graphic was created in Inkscape (internet freeware). The starting point for each graphic was the current IM display, for example, the entry of Ownship information shown in Figure 4. That graphic was then modified in Inkscape based on discussions by the interns and the IM team to meet the objectives outlined in paragraph 2.1 (address previously identified shortfalls, comply with new requirements, support new IM clearance types, etc.).

Once the modified Inkscape graphic was complete, it was embedded into a Power Point slide presentation, and then hyperlinks were added to provide the tool to with a limited emulation of data entry required for the IM system. In addition to data entry and progressions between those pages, additional hyperlinks were also added to the prototype tool to provide the transitions between ASTAR states (within the limitations of using hyperlinks). A slide from the prototype tool is shown in figure 2. A series of slides at the end of the tool illustrated what the corresponding CGD will look like (not shown).

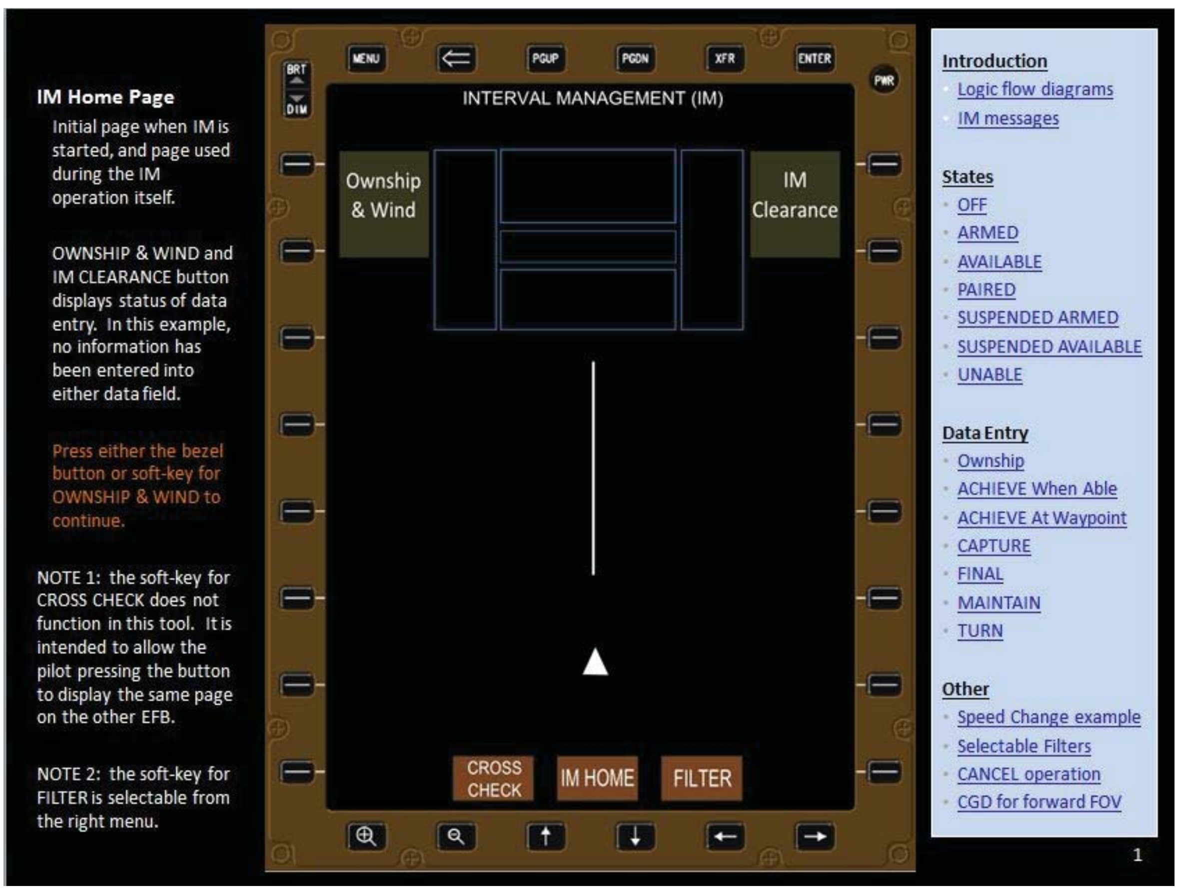

Figure 2. Revised IM display as created by the prototype tool.

Mental tracking of data entry had been noted during previous research as an issue within the current system because of the many different pages users have to view when entering data. One of the objectives for the prototype display was to decrease the amount of software pages required, while also enabling the user to see all the required fields of data that had to be entered on one page.

Similar to EAGAR, the prototype tool only supported one airport, the Phoenix Sky Harbor International Airport in Phoenix, Arizona (KPHX). Unlike the EAGAR tool, the prototype tool did support all five IM clearance types described in reference 10. Once data entry was complete, there was no connectivity or other software that generated information required to calculate and display the IM speed.

The overall layout was changed to make data entry easier to follow and understand. The IM data entry process has two sections: ownship and IM clearance. The ownship data entered by the pilot provides information about the aircraft they are flying. The IM clearance data entered by the pilot is information issued by ATC that is derived from the advanced arrival schedule software. (The specific information entered by the subjects for each tool is listed in the Protocol section.) Figure 3 shows an example of the current and prototype displays after all the ownship and IM clearance data had been entered. Due to limitations of the two tools, this study was not able to use the same airport, nor explore all the different spacing algorithm states.

As illustrated in figure 3, there are two key differences between the current display and the prototype tool on the IM home page. First, both the ownship and IM clearance information have been consolidated from three separate boxes into one box, and that box is now labeled. Second, at the bottom of the prototype IM home page (right side of figure 3), new soft keys are provided to enable the tool to comply with requirements listed in reference 10.

Figure 3. IM home page for current (left) and prototype (right) display.

2.3. Study Design

2.3.1. Independent Variable

The independent variable in this study was the IM display, with the two tools providing each display type. The current IM system used the scratchpad method for data entry where information is entered before specifying what data field the information will populate within the software. The prototype display used a data entry method where the data field to enter information into is chosen first, then the information is entered. Figure 4 shows an example of the two data entry methods. On the left, the EAGAR tool shows that the letter “K” has been entered into the scratchpad area of the display, whereas in the prototype tool, the letter “K” has been entered into the airport data field.

Figure 4. Ownship data entry for current (left) and prototype (right) display.

2.3.2. Test Matrix

The study compared the two display interfaces; therefore a (1 x 2) matrix was used. Each subject entered the ownship and IM clearance information into both displays. To account for and minimize the human learning effect, the order in which the displays were presented to the participants during data collection was randomized. However, for training consistency all subjects were taught the current display first, then the prototype tool.

In addition to the independent variable of IM display type, the study participants, who were IM subject matter experts, were also asked to provide qualitative comments about other aspects of the prototype tool.

2.3.3. Metrics

Both quantitative and qualitative metrics were used in this study. Quantitative measurements were taken for the time required for ownship data entry, IM clearance data entry, and the number of errors or confusion events. (An “error” or “confusion event” was defined as the occasion when a subject had an extended pause, was unable to find a particular button, frequently pushed a wrong button, had to go back to a previous display page, required assistance from the researcher, or attempted to enter incorrect information.)

Qualitative metrics were based on comparative ratings of intuitiveness and ability to mentally track data entry. Within the questionnaire there was also a section for comments where each participant had the opportunity to provide constructive criticism and recommendations for any changes that could help in improving the two displays. The comments section provided some of the most useful information in regards to further improving the IM display.

2.4. Subject Participants

Every participant in the study was a NASA employee or summer intern, and age, race, gender, and ethnic background were not factors to qualify as a participant. The volunteer participants for the study were categorized as either subject matter experts (SME) or non-SMEs, based on their knowledge of the IM system and procedures.

SME participants were expected to evaluate both the intuitiveness and the ease of use of the prototype tool, and to evaluate difficulties in comprehension, identify any issues with data entry into the prototype tool, as well as recommend changes to the tool. They were also instructed to assess whether the ownship and IM clearance data entries were valid, efficient, and logical, and they were asked to evaluate the additional features of the revised display (for example, the pilot-selectable filters and full map mode).

Non-SME participants were instructed to evaluate only the part of the questionnaire that focused on the intuitiveness and efficiency of data entry into the display, and to provide feedback in regards to how smoothly data was entered in the displays. It was not expected that they understand nor try to interpret the information entered, nor were they asked to evaluate the additional features added to the prototype. The expectation was that they only focus on how well the prototype tool was organized.

There were eight SMEs and fifteen non-SMEs who volunteered to participant in the study. The average time to complete the study was approximately one hour and thirty minutes for the SMEs, and approximately thirty minutes for the non-SMEs. Note that some of the SME participants participated from a remote location, therefore some of the tabular data presented later indicate six SMEs for quantitative metrics since that was not accomplished by the researcher (e.g., data entry time), while other quantitative metrics indicate eight SMEs participated for those items that were accomplished (e.g., ratings for intuitiveness and mental tracking).

2.5. Protocol

Prior to data collection, the participants given a general overview of the study, signed an Informed Consent form, and then received specific instructions and training prior to commencing data collection and completing the questionnaire.

During training, each participant was shown how to enter the information into the current display (the EAGAR tool), and then given the opportunity to practice as much as needed in order to achieve proficiency. The participant was then given training on the prototype tool (the Power Point slides with hyperlinks), and then, again, given the opportunity to practice as much as needed to achieve proficiency. They were informed that once data collection began, no assistance would be provided to them by the researcher as they entered the data into the two tools. At the end of the training for entering data into the tools, the questionnaire was reviewed and instructions given on how to complete it.

Data collection immediately followed the training session; each participant was given a document with the ownship and wind information, and the IM clearance for a particular display (the order of which was randomized). He or she then entered this information in the appropriate display while the researchers recorded quantitative data, including the entry times and confusion events. The volunteer was then given the information for the other display type, and, again, the researchers recorded quantitative data. Once the volunteer subject had completed both display types, they completed the questionnaire (for the SMEs the entire questionnaire, and for the non-SMEs just a portion of the questionnaire). Due to the limitations of time and resources, the data entered during the training session was identical to data entered during data collection.

Current Display

Ownship Data

- Cruise Alt FL360
- Cruise Mach .80
- Descent M/CAS .80/270
- Airport KMWH
- Route EAGAR1
- Transition SINGG
- Approach RRZ32R
- Descent Wind 1000’ 060/12

IM Clearance Data

- Goal 86
- Target SWA1756
- Route EAGAR1
- Transition EPENE
- Approach RRZ32R

Prototype Display

Ownship Data

- Airport KPHX
- Route MAIER5
- Transition BLD
- Approach ILS08
- Transition* JAMIL
- Surface Wind 060 / 012
- Surface Temp 15

IM Clearance Data

- Type ACHIEVE
- Initiate When Able
- Spacing 78
- Target DAL3267
- Route EAGUL5
- Approach ILS08
- Achieve Pt WAZUP
- Terminate Pt WAZUP

At the end of the study session, each participant then completed a paper questionnaire to rate the intuitiveness and ease of mentally tracking the data entry process. Only the SME participants were also asked to rate the usefulness of various data elements on the EFB and CGD displays.

3. Results

3.1. Entry Times

The subjects were asked to enter the ownship and IM clearance data into both displays while the researchers recorded the time it took to complete each portion to the nearest hundredth of a second. For each ownship and IM clearance entry, the difference between the current display entry time and the prototype tool entry time was computed. That difference was then divided by the current display entry time in order to calculate the percent difference of the prototype tool entry time from the current display entry time. Thus, if the percent difference was positive, the entry time for the prototype was faster than the entry time for the current system. If the percent difference was negative, the entry time for the prototype was slower than the entry time for the current system.

The percent differences were then tested to determine if there was a significant difference between SME results and the non-SME results for each entry type (ownship and IM clearance). If there was no difference between the results for the subject types, then the data for the two groups would be consolidated into one data set to test if the percent difference was operationally significant. If there was a difference between the groups, then each group would be tested separately on that display for operational significance. The research team established a mean difference greater than 10% as operationally significant. Table 1 shows descriptive statistics of the entry times for the study.

Table 1. Entry times (in seconds) by display type by subject type

Data Entry Type	Subject Type	Display	*N*	Mean	Standard Dev	Min	Median	Max
Ownship and Wind	SME	Current	6	68.49	9.85	55.00	67.00	85.00
		Prototype	6	60.94	14.23	42.62	58.50	80.00
	Non-SME	Current	15	97.97	25.60	53.03	89.62	146.55
		Prototype	15	54.26	18.55	29.50	55.00	108.98
IM Clearance	SME	Current	6	38.96	7.57	32.00	37.38	50.00
		Prototype	6	39.15	8.71	30.00	35.52	53.88
	Non-SME	Current	15	46.23	14.48	32.89	40.41	79.43
		Prototype	15	34.97	13.36	19.57	31.03	72.23

3.1.1. Determining Differences in Data Entry Time by Group

A two-sample *t*-test was conducted to determine if there was a difference in the mean data entry times for the SME and non-SME participants (ref. 12). Table 2 below shows the *p*-values for the tests on the two-entry types, as well as the decision on whether or not to reject the null hypothesis. The null hypothesis was rejected when the *p*-value < α (0.05).

Table 2. *T*-test results for ownship and clearance entry time differences

Entry Type	*p*-value	95% Confidence Interval	Decision
Ownship and Wind (SME ≠ Non-SME)	0.020	(6.8%, 56.7%) (Non-SME minus SME)	Reject the null hypothesis that there is no difference between the percent differences.
IM Clearance (SME ≠ Non-SME)	0.082	(–3.5%, 50.5%) (Non-SME minus SME)	Fail to reject the null hypothesis that there is no difference between the percent differences.

At the 0.05 level of significance, there is enough evidence to conclude that the mean percent difference for SME is different than that for non-SME for the ownship entry times ($p = 0.020$). There is 95% confidence that the true mean difference between SME and non-SME is within the interval (6.8%, 56.7%). Since the mean is different for this entry time, the two groups were tested separately to assess whether the mean percent difference is operationally significant. There is not enough evidence to conclude that the mean percent difference for SME is different than the mean percent difference for non-SME for the clearance entry times ($p = 0.082$), therefore the data from the two subject groups for the IM clearance data entry was combined.

3.1.2. Analysis of Data Entry Time by Type

Ownship data entry percent differences were analyzed separately for SME and non-SME groups, and IM clearance data entry for the two groups was combined for analysis (explained above). One-sample *t*-tests were conducted on the percent differences for SME ownship entry, non-SME ownship entry, and all-subject clearance entry to determine if each was greater than 10% at the $\alpha = 0.05$ level of significance (ref. 12). Table 3 shows the sample means, *p*-values, and conclusions for the tests.

Table 3. Means, standard deviations, and *p*-values for percent differences

Null Hypothesis	*N*	Mean	Standard Deviation	*p*-value	Decision
The difference in time needed for Ownship and Wind data entry by SME subjects is ≤ 10% between the displays types	6	9.71%	23.33%	0.512	Fail to reject the null hypothesis
The difference in time needed for Ownship and Wind data entry by non-SME subjects is ≤ 10% between the displays types	15	41.46%	17.67%	< 0.0005	Reject the null hypothesis
The difference in time needed for IM clearance data entry for either group of subjects is ≤ 10% between the displays types	21	13.93%	29.30%	0.273	Fail to reject the null hypothesis

At the 0.05 level of significance, there is not enough evidence to conclude that the mean percent difference for SME ownship entry times is greater than 10% ($p = 0.512$), therefore it cannot be concluded that the SME ownship entry was 10% faster using the prototype tool than the current tool. However, there is enough evidence to conclude that the mean percent difference for non-SME ownship entry times is greater than 10% ($p < 0.0005$); therefore non-SME ownship data entry was more than 10% faster on the prototype tool than on the current tool. Furthermore, there is 95% confidence that the true mean percent difference for non-SME ownship entry is greater than 33.42%. Finally, there is not enough evidence to conclude that the mean percent difference for IM clearance entry times is greater than 10% ($p = 0.273$) for either group of subjects.

3.2. Intuitiveness and Mental Tracking Ratings

All subjects were asked to complete a questionnaire item rating each display using the seven-point Likert rating scale, seen in table 4, from "1" (completely disagree) to "7" (completely agree) intended to assess the intuitiveness and ease of use of the displays. Table 5 shows descriptive statistics for the ratings by the groups of subjects on the different entry types for the different displays.

Table 4. Questionnaire for intuitiveness and mental tracking of data entry

	Rating Scale						
	Completely Disagree			**Neutral**			**Completely Agree**
Intuitive design of data entry:							
#1	1	2	3	4	5	6	7
#2	1	2	3	4	5	6	7
Mentally track data entry progress:							
#1	1	2	3	4	5	6	7
#2	1	2	3	4	5	6	7

Table 5. Results for intuitiveness and mental tracking for both displays by subject

Rating	Subject Type	Display	*N*	Mean	Standard Deviation	Min	Median	Max
Intuitiveness	SME	Current	8	5.50	0.93	4.0	5.5	7.0
		Prototype	8	5.25	0.89	4.0	5.5	6.0
	Non-SME	Current	15	3.67	1.29	2.0	4.0	5.0
		Prototype	15	5.60	0.63	4.0	6.0	6.0
Mental Tracking	SME	Current	8	5.50	0.76	4.0	6.0	6.0
		Prototype	8	5.88	0.35	5.0	6.0	6.0
	Non-SME	Current	15	3.73	1.16	2.0	3.0	6.0
		Prototype	15	5.47	1.13	3.0	6.0	7.0

3.2.1. Differences Between SME and Non-SME Ratings

Prior to comparing the current tool with the prototype tool, the data were tested to see if there were differences between the SME and non-SME subject groups. A Wilcoxon Mann-Whitney Rank Sum Test of the differences between the SME and non-SME ratings was conducted (ref. 12). Figure 5 shows boxplots comparing the ratings for both displays by subject type, and table 6 shows the *p*-values for the tests of SME ratings vs. non-SME ratings as well as the conclusions from the tests.

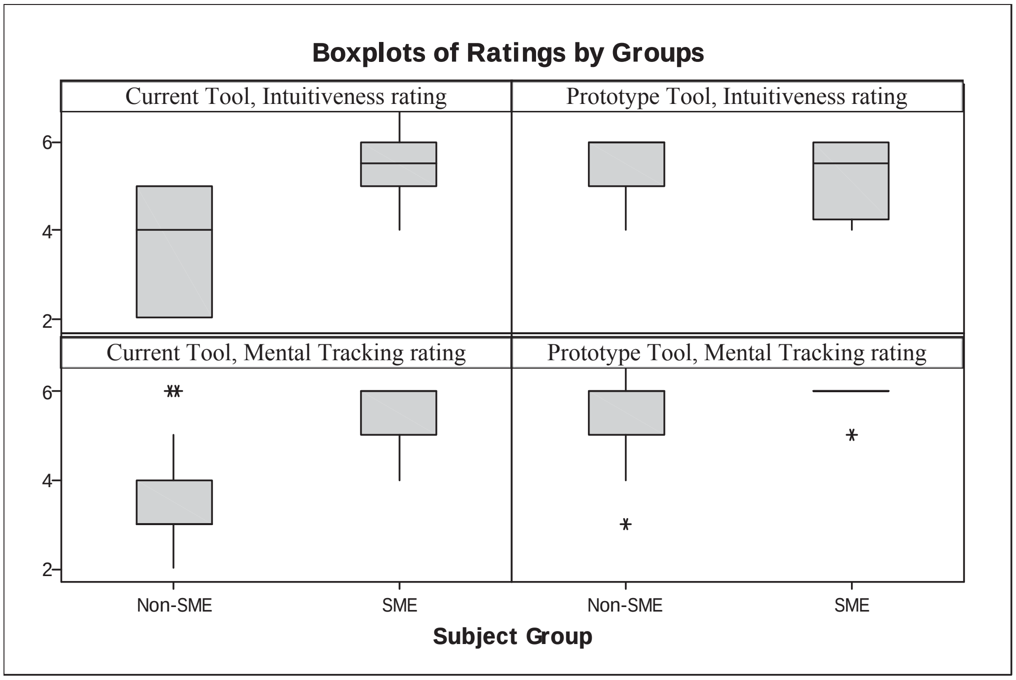

Figure 5. Boxplots of intuitiveness and mental tracking ratings grouped by subject type.

Table 6. *P*-values and conclusions for intuitiveness and mental tracking by subject

Rating Comparison	*p*-value	Decision
Current Display Intuitiveness (SME ≠ non-SME)	0.0050	Reject the null hypothesis that there is no difference between the ratings by the two groups for the current display.
Prototype Tool Intuitiveness (SME ≠ non-SME)	0.4197	Fail to reject the null hypothesis that there is no difference between the ratings by the two groups for the prototype tool.
Current Display Mental Tracking (SME ≠ non-SME)	0.0142	Reject the null hypothesis that there is no difference between the ratings by the two groups for the current display.
Prototype Tool Mental Tracking (SME ≠ non-SME)	0.6128	Fail to reject the null hypothesis that there is no difference between the ratings by the two groups for the prototype tool.

At the 0.05 significance level, there is enough evidence to conclude that the SME rating of the current display is different than that for non-SME for both the intuitiveness ($p = 0.0050$) and the mental tracking of data entry ratings ($p = 0.142$). In fact, the SME participants rated the current display significantly higher than the non-SMEs (see figure 5). There is not enough evidence to conclude that the median rating of the prototype tool by SME is different than that for non-SME for either the intuitiveness ($p = 0.4197$) or the mental tracking of data entry ratings ($p = 0.6128$). Due to the differences with the current display, subject groups were analyzed separately.

3.2.2. Comparative Intuitiveness Rating

Since the intuitiveness ratings were statistically different for SME and non-SME participants, the ratings were analyzed by subject group using a Wilcoxon Mann-Whitney Rank Sum Test to determine if the prototype tool received higher median intuitiveness ratings than the current display tool (ref. 12). Figure 6 shows boxplots of the intuitiveness ratings for the non-SME subject group, Figure 7 shows boxplots of the intuitiveness ratings for the SME subject group, and table 7 shows the *p*-values and conclusions.

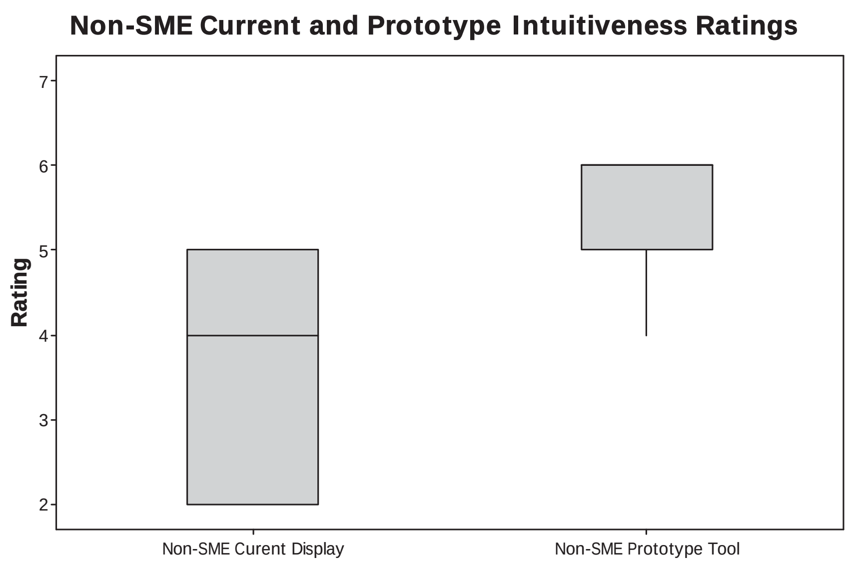

Figure 6. Intuitiveness boxplots by display type by non-SME group.

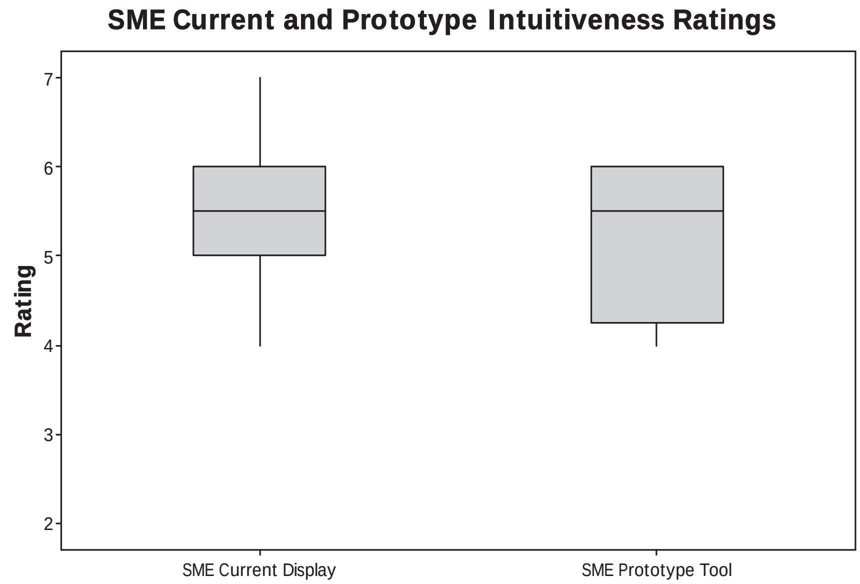

Figure 7. Intuitiveness boxplots by display type by SME group.

Table 7. *P*-values and conclusions for intuitiveness by display type by subject group

Hypothesis	*p*-value	Conclusion
SME Prototype > SME Current Display Intuitiveness Rating	0.6773	Fail to reject the null hypothesis that the SME mean intuitiveness rating for the proposed display is equal to that of the current display.
Non-SME Prototype > Non-SME Current Display Intuitiveness Rating	0.0001	Reject the null hypothesis that the non-SME mean intuitiveness rating for the proposed display is equal to that of the current display.

There is not enough evidence to conclude that the median intuitiveness rating for the prototype tool is greater than the median intuitiveness rating for the current display by the SME ($p = 0.6773$). However, the non-SMEs rated the prototype tool as more intuitive than the current display ($p = 0.0001$).

3.2.3. Comparative Mental Tracking Rating

Since the mental tracking ratings were statistically different for SME and non-SME participants, the ratings were analyzed by subject group using a Wilcoxon Mann-Whitney Rank Sum Test to determine if the prototype tool received higher median ratings than the current display tool for ease of use (ref. 12). Figure 8 shows boxplots of the mental tracking ratings for the non-SME subject group, Figure 9 shows boxplots of the mental tracking ratings for the SME subject group, and table 8 shows the *p*-values and conclusions.

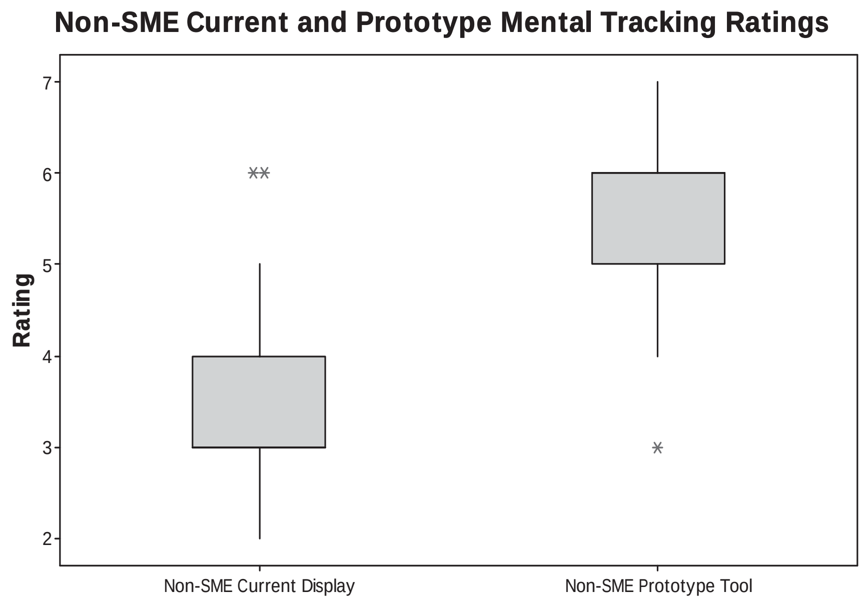

Figure 8. Boxplots of mental tracking ratings by display type by non-SME subject group.

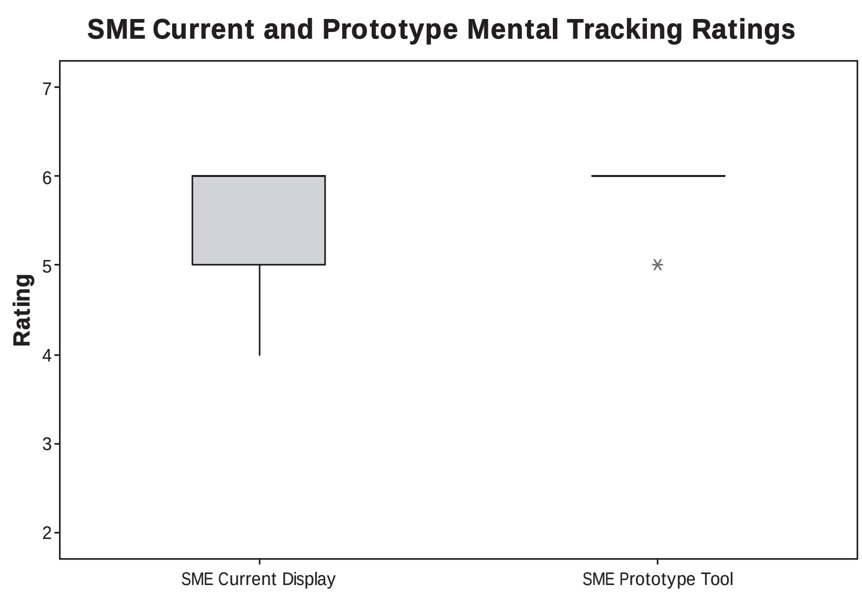

Figure 9. Boxplots of mental tracking ratings by display type by SME subject group.

Table 8. *P*-values and conclusions for mental tracking by display type by subject group

Hypothesis	*p*-value	Conclusion
SME Prototype > SME Current Display Mental Tracking Rating	0.2004	Fail to reject the null hypothesis that the SME prototype mental tracking rating is equal to the median current display mental tracking rating.
Non-SME Prototype > Non-SME Current Display Mental Tracking Rating	0.0007	Reject the null hypothesis that the SME prototype mental tracking rating is equal to the median current display mental tracking rating.

At the 0.05 level of significance there is not enough evidence to conclude that the SME mental tracking rating for the prototype tool is greater than for the current display ($p = 0.2004$). However, the non-SME participants found it easier to mentally track data entry with the prototype tool than with the current display ($p = 0.0007$).

3.2.4. Intuitive Sufficiency

Regardless of whether or not the prototype tool is rated as better than the current display, to be operationally sufficient the prototype tool should exceed a median rating of 5 for intuitiveness. A Wilcoxon Signed Rank Test was performed (ref. 12); figure 10 presents a boxplot of the intuitiveness ratings for all subjects, and table 9 presents the *p*-values and conclusions for the hypothesis test. There is enough evidence to conclude that the median intuitiveness rating for the prototype tool is greater than five, so the intuitiveness rating is operationally sufficient.

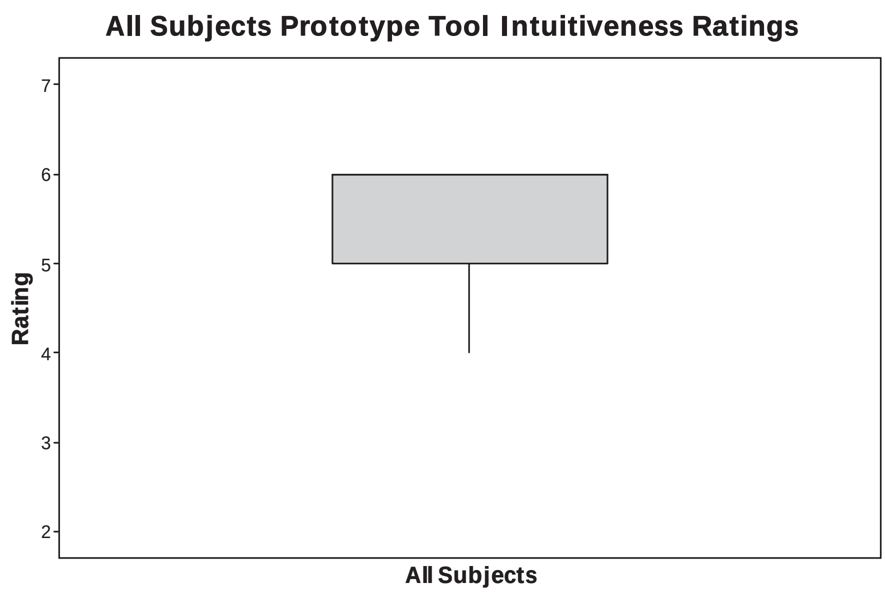

Figure 10. Boxplot of intuitiveness for all subjects.

Table 9. *P*-value and conclusion of intuitiveness rating for all subjects

Hypothesis	*N*	Median	*p*-value	Conclusion
Prototype Tool Intuitiveness Rating > 5	23	6.0	0.010	Reject the null hypothesis that the mean intuitiveness rating for the prototype tool is equal to or less than five.

3.2.5. Mental Tracking Sufficiency

Regardless of whether or not the prototype tool is rated as better than the current display, to be operationally sufficient the prototype tool should exceed a median rating of 5 for mental tracking. A Wilcoxon Signed Rank Test was performed (ref. 12); figure 11 presents a boxplot of the mental tracking ratings for all subjects, and table 10 shows the *p*-values and conclusions for the hypothesis test. There is enough evidence to conclude with 95% confidence that the median mental tracking rating for the prototype tool is greater than five, indicating that it is operationally sufficient.

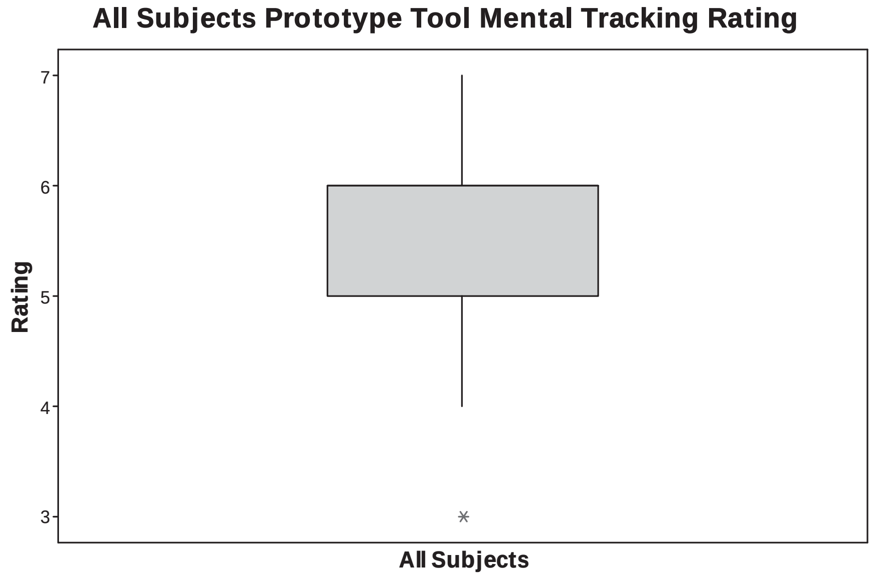

Figure 11. Boxplots of mental tracking ratings for all subjects.

Table 10. *P*-value and conclusion for mental tracking for all subjects

Hypothesis	*N*	Median	*p*-value	Conclusion
Prototype Tool Mental Tracking Rating > 5	23	6.0	0.008	Reject the null hypothesis that the median intuitiveness rating for the prototype tool is equal to or less than five.

3.3. Number and Type of Errors

While the participants were entering the ownship and clearance data, the researchers recorded any errors or confusion encountered by the subjects. Errors were grouped by data entry type (ownship or IM clearance) and by display type (current or prototype). The most common errors that occurred when subjects used the prototype tool were addressed by additional changes to a revised version of the prototype tool that was delivered to the software development teams (some of specific changes listed in the Conclusion section of this paper).

For the current display, there were 11 confusion events for the ownship entry which varied in their causes. The most frequent error was that four participants had trouble finding the “ENTER” button after completing ownship entry, which is required to enter the forecast descent wind information. Figure 12 is a screen capture from the HyperCam2 file for a particular subject who needed approximately 15 seconds to find the “ENTER” button to be able to enter the wind information. For the current display IM clearance entry, there were only three errors and they were for repeatedly pressing the manual entry button and having to re-enter the target aircraft ID. Two of the non-SME subjects made this error more than twice on the same run.

Figure 12. Image capture of data entry error in current display tool.

For the prototype tool, the errors that occurred generally involved loading the wind forecast for the ownship entry (nine of the fourteen prototype ownship errors) and finding the “ENTER” button after entering the target ID for the IM clearance entry (six of the seven prototype clearance confusions). Figure 13 is a screen capture from the HyperCam2 file for a particular subject who needed approximately 8 seconds to find the “LOAD WIND FORECAST” button to be able to enter the wind information. It was reported by the subjects that all data entry prior to this point had been a sequential flow from top-left to bottom-left of the EFB, therefore they had not expected the next button push to be on the right side of the EFB.

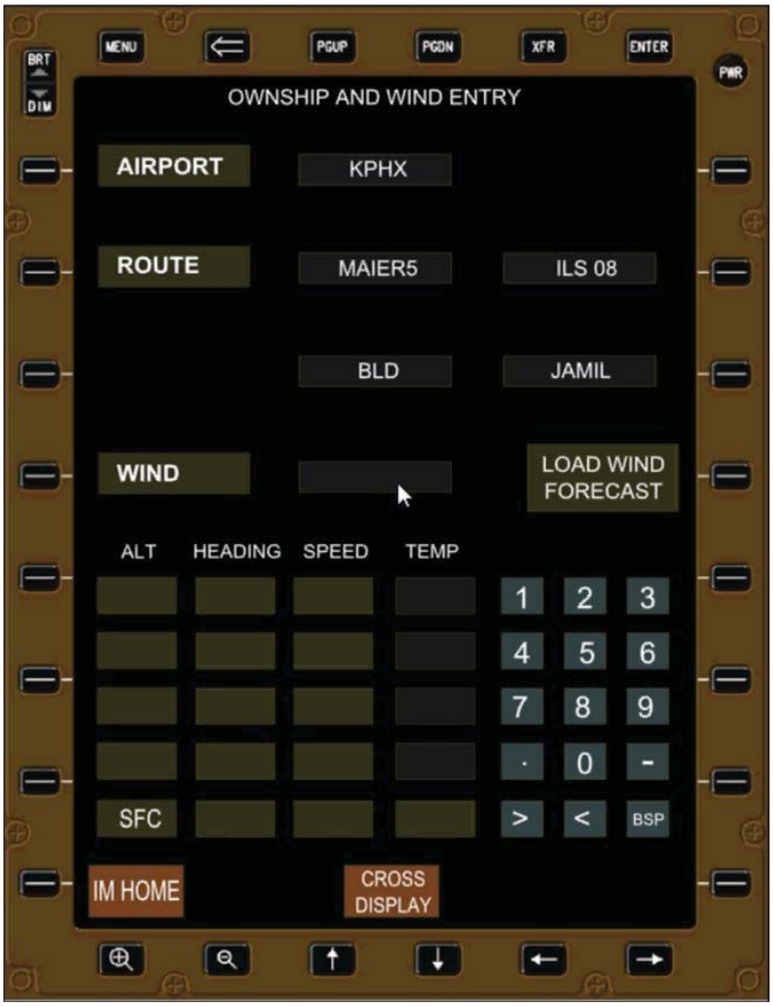

Figure 13. Image capture of wind entry error in prototype display tool.

Only one subject made more than one type of error for ownship data entry for either display type, and no subject made more than one type of error on the IM clearance data entry for either display. All but one of the SME subjects made an error using the prototype tool ownship entry accounting for six of the 14 errors on this section, four of which were errors with the forecast winds entry (illustrated in figure 13). Table 11 and Table 12 show the number of errors made by data entry type and by subject type.

Table 11. Number of errors by category for current display

	Ownship					Clearance
	Clicking "ENTER" After Completed Ownship	Wind Entry	Entering Ownship Route	Entering Approach	Entering Destination Airport	Pressing "ENTER" After Target ID
SME	0	2	0	0	0	0
Non-SME	4*	1	2	1	1	3
Total	4	3	2	1	1	3

Note: "*" indicates one of them is shown in figure 12.

Table 12. Number of errors by category for prototype display

	Ownship				Clearance	
	Forecast Winds	Selecting the Field or Bezel Button	Page Down	Interval Management	Target Aircraft Enter	Enter After Target Route
SME	4	1	0	0	2	0
Non-SME	5	0	1	1	4	1
Total	9	1	1	1	6	1

3.4. Data Elements Ratings

Only SME participants were asked to rate the usefulness of additional data elements using a 7-point scale, where "1" corresponds to "not very useful" and "7" corresponds to "very useful." The elements are displayed in two different areas. The first is the CGD which will be in the pilot's primary field of view (FOV), and the second is the EFB which is outside the pilot's FOV (ref. 13). Some of the SME's did not rate all of the data elements on both of the display types because they felt that their level of expertise was not adequate to answer certain items. Table 13 displays descriptive statistics for the usefulness ratings of the data elements for the CGD and the EFB. Based on the subjective judgment of the research team, means greater than six are highlighted in green (indicating high usefulness), means between five and six are highlighted in yellow (indicating somewhat useful), and means less than 3.5 are highlighted in red (indicating not useful or desired).

Table 13. Usefulness rating by data elements by device by subject matter experts

Data Element	*N*	Mean	Standard Deviation	Min	Median	Max
CGD Ownship	6	4.170	2.48	1	4	7
CGD Clearance	6	4.500	2.258	2	4.5	7
CGD FIM Speed	6	7.000	0.0	7	7	7
CGD FIM Status	6	6.667	0.516	6	7	7
CGD FIM Message	6	6.500	0.837	5	7	7
CGD Fast/Slow	6	5.833	1.329	4	6	7
CGD Early/Late	6	5.667	1.506	3	6	7
CGD Bearing, Range, Altitude	6	2.833	1.329	1	3	5
CGD Ground Speed	6	3.333	1.366	1	3.5	5
CGD TRK	6	2.667	1.366	1	2.5	5
EFB Ownship	7	5.714	1.799	2	6	7
EFB Clearance	7	6.000	1.915	2	7	7
EFB FIM Speed	7	6.571	0.787	5	7	7
EFB FIM Status	7	6.571	0.535	6	7	7
EFB FIM Message	7	6.857	0.378	6	7	7
EFB Fast/Slow	7	4.714	0.756	4	5	6
EFB Early/Late	7	5.286	0.951	4	6	6
EFB Bearing, Range, Altitude	7	4.429	2.070	1	5	6
EFB Ground Speed	7	3.714	2.059	1	4	6
EFB TRK	7	4.000	1.915	1	4	6

The information in Table 13 is shown again in Figure 14 as a boxplot of the ratings by data element on the CGD, and in Figure 15 as a boxplot of the ratings by data element on the EFB.

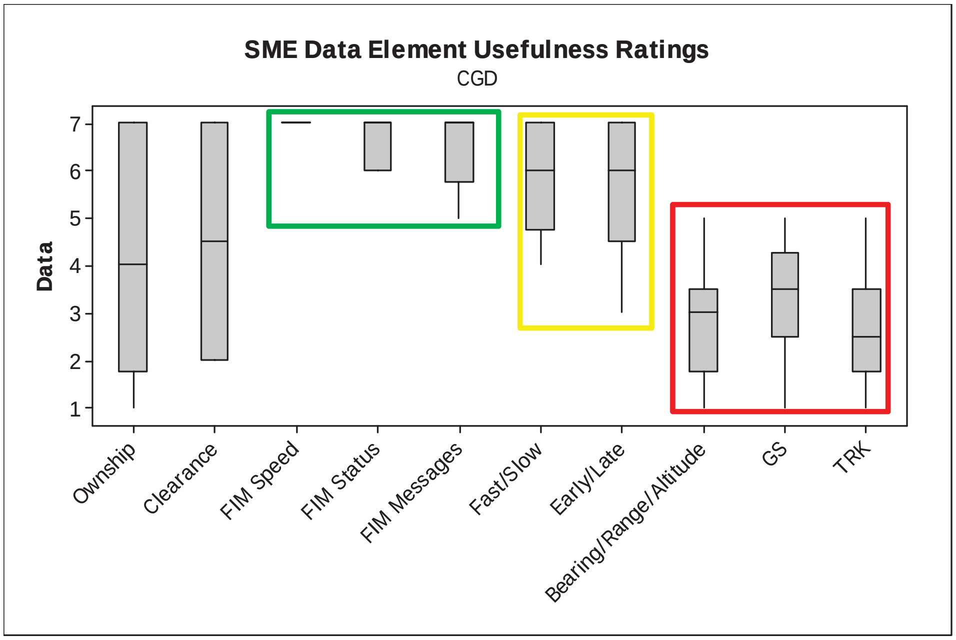

Figure 14. Boxplots of data element usefulness ratings on CGD.

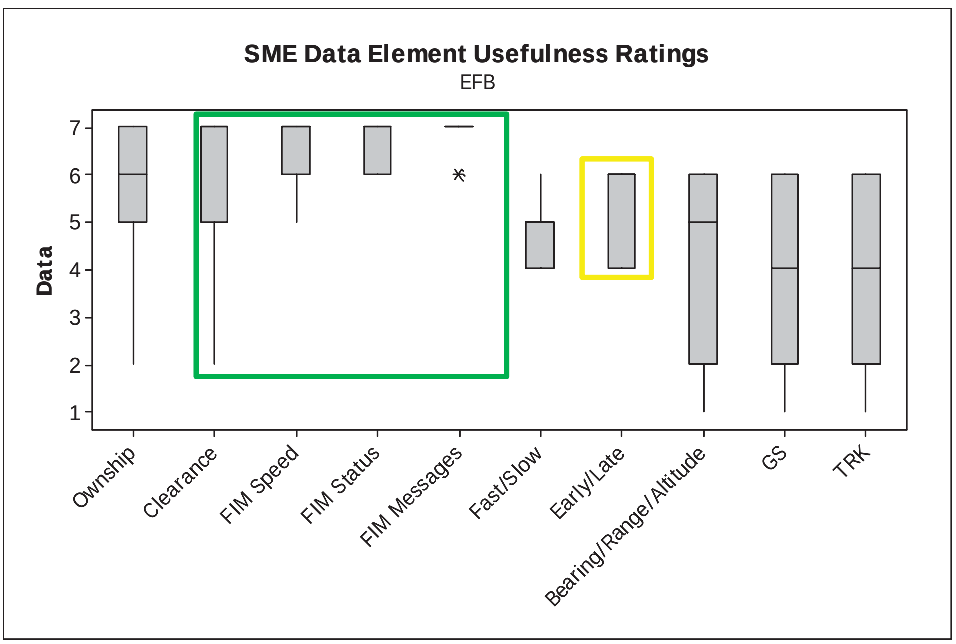

Figure 15. Boxplots of data element usefulness ratings on EFB.

4. Conclusion

For the qualitative metrics, the time for the SME participants to enter data into the prototype tool and the current tool was the same, whereas the non-SME participants were significantly faster entering data into the prototype compared to the current tool. The research team hypothesizes the lack of difference of the SME participants is due to their extensive knowledge and familiarity with the current tool, which many of them either helped create or use as part of their daily tasks. Neither subject group entered the IM clearance data 10% faster in the prototype tool compared to the current display tool.

For the ratings of intuitiveness and mental tracking, there was again a difference between how the two subject groups rated the displays. The SME participants did not rate the intuitiveness or the mental tracking required to be any different between the current and prototype displays, whereas the non-SME participants did rate the prototype as an improvement over the current display tool. The research team again hypothesizes the SME's familiarity and daily use of the current display tool may have impacted the rating they gave.

Both SME and non-SME participants rated the intuitiveness of data entry and the ability to mentally track the progress of data entry in the prototype tool as greater than "5" on a scale of "1" (completely disagree) to "7" (completely agree). The research team interpreted that to mean that the prototype tool is sufficient for operational use, regardless of the rating when compared to the current tool.

Errors and confusion by SME participants when using the current display tool where almost non-existent, while non-SME participants showed difficulty pressing the "ENTER" button to enable entering the forecast descent wind information, and pressing the "ENTER" button after entering the target aircraft's identification. Errors and confusion by the SME participants when using the prototype display tool was predominately caused by the "LOAD FORECAST WIND" button being located on the right side of the EFB, breaking the linear progression of data entry they were accustomed to in the prototype tool. Non-SME participants using the prototype tool also had challenges with the "LOAD FORECAST WIND" button, as well as completing the entry of the target aircraft identification.

Only the SME participants rated the data elements located on the CGD (primary FOV) and EFB (outside of primary FOV). Elements such as the IM speed, IM status, and IM messages received high ratings of usefulness on both the CGD and EFB. The elements of target bearing, range, altitude, ground speed, and ground track, when located on the CGD, received very low ratings of usefulness.

This study was completed in time for the interns to use the results to revise the prototype tool and documentation prior to delivering it to the software development team. A partial list of improvements made based on this study follows:

- All data fields are now only accessible by bezel button or soft-key (data field itself removed as an option).
- The ownship and target route are reduced to one data field row.
- The wind data field states either "EMPTY" or provides a time stamp of when the wind message was sent.
- The "LOAD FORECAST WIND" button is placed directly below the wind field.

- The ability to manually enter the target identification from the IM clearance home page was removed, and must be entered from the target ID page.
- The location to select manual entry on the target ID page was raised above the bottom row.
- The keyboard was compressed (by removing the "ENTER" button and shrinking the gap between rows), allowing the target route data field to remain visible when the keyboard is present.
- The "IM home" was modified to be at the bottom-center of the EFB, and will change color to indicate when pressing it will have no effect (i.e., already on the IM home page).
- The logic to transition between different IM states was refined.
- A "confirm cancel IM clearance" page was added.

In summary, for the SME participants, the prototype tool did not appear to provide a clearly improved set of displays in terms of time to enter data, intuitiveness, ability to track the progress of data entry, or the number of errors when entering data, for either ownship or IM clearance data types. However for the non-SME participants, the prototype tool did appear to consistently rate better, in some cases significantly better than the current displays.

The prototype tool does, however, provide IM displays which address some of the issues raise in previous research, and meets almost all of the new requirements recently given to the research team, which the current IM displays do not. Finally, this prototype tool has been delivered to the software development teams at NASA Langley, and provides the basis for the displays to be used in the large IMAC human-in-the-loop experiment to be conducted in the summer of 2015.

5. References

1. *Federal Aviation Administration, FAA Aerospace Forecast*, Fiscal Years 2013–2033.
2. Baxley, B., et al., *Air Traffic Management Technology Demonstration-1 Concept of Operations (ATD-1 ConOps), Version 2.0*, NASA/TM 2013-218040, Sept 2013.
3. Baxley, B., Murdoch, J., Swieringa, K., Barmore, B., Capron, W., Hubbs, C., Shay, R., Abbott, T., *Experiment Description and Results for Arrival Operations Using Interval Management with Spacing to Parallel Dependent Runways (IMSPiDR)*, NASA/TP-2013-217999, June 2013.
4. Swieringa, K., Murdoch, J., Baxley, B., Hubbs, C., *Evaulation of an Airborne Spacing Concept, On-board Spacing Tool, and Pilot Interface*, AIAA 2010.
5. Murdoch, J., Wilson, S., Hubbs, C., Smail, J., *Acceptability of Flight Deck-Based Interval Management Crew Procedures*, AIAA Technical Conferences 2013.
6. Wilson, S., Murdoch, J., Hubbs, C., Swieringa, K., *Evaluation of Flight Deck-Based Interval Management Crew Procedure Feasibility*, AIAA 2013.
7. Latorella, K., *Avionics Configuration Assessment for Flightdeck Interval Management: A Comparison of Avionics and Notification Methods*, NASA/TP in-press.
8. Swieringa, K., Wilson, S., Shay, R., *An Investigation of Interval Management Displays*, NASA/TM-2014-217743, May 2014.
9. Swieringa, K., Wislon, S., Shay, R., *An Evaluation of Retrofit Flight Deck Displays for Interval Management*, AIAA ATIO 2014.
10. *Safety, Performance and Interoperability Requirements Document for Airborne Spacing—Flight Deck Interval Management (ASPA-FIM), RTCA Paper 001-11 / SC-186-330, draft* V1.0, May 19, 2011.
11. Baxley, B., Shay, R., Swieringa, K., *The Development of Cockpit Display and Alerting Concepts for Interval Management (IM) in a Near-Term Environment*, NASA/TM-2014-218659, November 2014.
12. Cabillo, P., Masaro, J., *Basic Statistical Procedures and Tables*, Acadia University, Tenth Edition, 2001.
13. *Federal Aviation Administration, Advisory Circular AC 25-11A, Electronic Flight Deck Displays,* June 2007.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01- 12 - 2014	Technical Memorandum	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Interval Management Display Design Study	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
Baxley, Brian T.; Beyer, Timothy M.; Cooke, Stuart D.; Grant, Karlus A.	5e. TASK NUMBER
	5f. WORK UNIT NUMBER 305295.02.41.07.07

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
NASA Langley Research Center Hampton, VA 23681-2199	L-20484
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
National Aeronautics and Space Administration Washington, DC 20546-0001	NASA
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) NASA/TM-2014-218661

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified - Unlimited
Subject Category 03
Availability: STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In 2012, the Federal Aviation Administration (FAA) estimated that U.S. commercial air carriers moved 736.7 million passengers over 822.3 billion revenue-passenger miles (ref. 1). The FAA also forecasts, in that same report, an average annual increase in passenger traffic of 2.2 percent per year for the next 20 years, which approximates to one-and-a-half times the number of today's aircraft operations and passengers by the year 2033. If airspace capacity and throughput remain unchanged, then flight delays will increase, particularly at those airports already operating near or at capacity. Therefore it is critical to create new and improved technologies, communications, and procedures to be used by air traffic controllers and pilots. National Aeronautics and Space Administration (NASA), the FAA, and the aviation industry are working together to improve the efficiency of the National Airspace System and the cost to operate in it in several ways, one of which is through the creation of the Next Generation Air Transportation System (NextGen). NextGen is intended to provide airspace users with more precise information about traffic, routing, and weather, as well as improve the control mechanisms within the air traffic system. NASA's Air Traffic Management Technology Demonstration-1 (ATD-1) Project is designed to contribute to the goals of NextGen, and accomplishes this by integrating three NASA technologies to enable fuel-efficient arrival operations into high-density airports (ref. 2).

15. SUBJECT TERMS

Air Traffic; Aircraft Industry; Airspace; Controllers; National Airspace System

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	35	(757) 864-9658