DOT/FAA/AM-04/20

Office of Aerospace Medicine
Washington, DC 20591

Usability and Effectiveness of Advanced General Aviation Cockpit Displays for Visual Flight Procedures

Kevin W. Williams
Jerry D. Ball
Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

December 2004

Final Report

This document is available to the public through the Defense Technical Information Center, Ft. Belvior, VA 22060, and the National Technical Information Service, Springfield, VA 22161.

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-04/20		

4. Title and Subtitle	5. Report Date
Usability and Effectiveness of Advanced General Aviation Cockpit Displays for Visual Flight Procedures	December 2004
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Williams KW, Ball JD	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125	11. Contract or Grant No.
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	14. Sponsoring Agency Code

15. Supplemental Notes

Work was accomplished under approved task AM-HRR-521.

16. Abstract

Twenty-four pilots participated in a study examining the usability and effectiveness of a set of advanced general aviation cockpit displays under visual flight procedures. Use of the displays for navigation and the ability to identify landmarks were measured using both objective and subjective formats. In addition, the effect of the displays on the ability to locate traffic out-the-window was measured both objectively and subjectively. Differences were found in the ability to navigate, to locate landmarks, and in the identification of traffic, when using the advanced displays. However, use of the displays decreased the amount of time that pilots spent looking outside of the aircraft. Implications of the findings are discussed.

17. Key Words	18. Distribution Statement
Human Factors, Advanced Cockpit Displays, Navigation, General Aviation, Usability, CDTI	Document is available to the public through the Defense Technical Information Center, Ft. Belvior, Va. 22060; and the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	21	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Usability and Effectiveness of Advanced General Aviation Cockpit Displays for Visual Flight Procedures

INTRODUCTION

As more advanced electronics make their way into the general aviation (GA) cockpit, there is a greater need to study the effects that these displays have on single-pilot operations. The inclusion of advanced displays in the GA cockpit has the potential to improve pilot situation awareness during critical portions of the flight. However, it also has the potential to distract the pilot from essential scanning activities and to increase workload by requiring pilot interaction during the flight.

We are aware of the negative effects that advanced cockpit displays can have on airline pilots (Sarter & Woods, 1994). These negative effects include the imposition of knowledge requirements, communication tasks, and attentional demands, as well as the potential for new forms of error or failure. Airline operations require two pilots in the cockpit. Pilots in single-pilot operations may be more prone to the negative effects imposed by advanced displays because there is no way to relieve any workload by offloading tasks to a second pilot.

A typical set of advanced navigation displays is shown in Figure 1. This particular configuration consists of a Global Positioning System (GPS) display and an accompanying multi-function display (MFD), and it is currently in use in a number of GA aircraft in Alaska as a part of the Alaska Capstone Program (FAA, 2004; Williams, 2002).

Most of the pilots participating in the Capstone Program are professional pilots, flying almost every day of the year. Consequently, the majority of these pilots fly over 100 hours each month, and nearly all (95%) are instrument rated. This leads to the obvious question of the impact of this kind of technology on the non-professional pilot population. The majority of non-professional pilots typically fly much less often, and a significant number do not have an instrument rating (FAA, 2000). In addition, the GA pilot population is more diverse, having a wider range of total flight experience and ages. What is needed is an examination of the effects of these displays on the non-professional GA pilot population.

Williams and Ball (2003) looked at the effects of advanced cockpit displays in single-pilot GA operations during instrument flight. That study compared the effect that the displays had during instrument flight on pilot performance and workload relative to standard navigation instruments. The study included a number of normal instrument flight procedures, including navigating toward a Very-High-Frequency Omnidirectional-Range (VOR) facility, conducting a holding procedure, and performing a missed approach. The study also examined the effect of high-workload conditions by introducing a vacuum-system failure during the flight. In addition to collecting pilot flight-technical performance and workload measures, the study also elicited subjective usability estimates from the participants.

Results of that study showed that pilots thought the displays were excellent for maintaining situation awareness during flight, although they believed that use of the displays might have had a negative effect on their scanning of aircraft instruments. Measures of workload did not show any differences between the displays. However, flight-performance measures showed superior flight performance using the advanced displays, under both normal instrument procedures and after the introduction of a failure of the vacuum system. Finally, the limited training provided to pilots was insufficient for them to master the new displays, as was demonstrated by the amount of assistance required to complete the display-interaction tasks and by the number of errors that occurred while completing those tasks.

The current study extends this initial research by examining the use of these displays during visual flight rules (VFR) flight. Again, the usability of the displays was measured, along with the effect of the displays on flight performance and workload. In addition, because the flight was in visual meteorological conditions (VMC), pilots were asked to locate and identify landmarks and traffic during the flight. In contrast to the previous research, the effect on pilot scanning was measured objectively using eye-tracking equipment.

METHODS

Participants

Twenty-four pilots were recruited from the Oklahoma City, OK metropolitan area and were compensated monetarily for their participation. All of the pilots were

Figure 1. Capstone Avionics displays. The MX-20 display is shown on top, and the GX-60 GPS unit is shown below it.

male, with an average age of 25.54 years, a median age of 23.5 years, and a range in ages from 20 to 42 years. Mean total flight hours was 840.46 hours with a median of 502.5 hours. The participants were asked to estimate the number of flight hours logged for several categories of flight. They were also asked whether they had used a cockpit GPS or multi-function display, if they had flown the aircraft model simulated in the experiment, how easy it was to learn to use the displays, and whether they would like to use these types of displays in actual flight. This information is summarized in the Results section.

Design

There were three display conditions in the study: conventional navigation displays (CON condition); use of a GPS display (GPS condition); and the use of both a GPS display and a moving-map display (MFD condition). Owing to the length of time to fly in a single condition and the need to avoid pilot fatigue, each pilot was only tested under two different display conditions. This resulted in a triple within-subject experimental design, corresponding to each of the possible pairwise pairings of the three display conditions (i.e., CON-GPS, CON-MFD, GPS-MFD).

Pilots were asked to plan and fly two separate flights in VMC. On each flight, a different display condition was specified. Selection of the display condition was determined by which display group the pilot was in. Within each display group, the order of display conditions and flight scenarios was counterbalanced across pilots. Pilots in all conditions were allowed use of a sectional chart for both flight planning and during the flight.

Measurement Instruments

Usability, workload, and pilot performance were gauged using a number of subjective and objective measures. A usability questionnaire was administered to each pilot to evaluate various aspects of the GPS (GX-50) and MFD (MX-20) displays. Four rating criteria (effectiveness, ease of use, cues and prompts, and system feedback) were scored using a 5-point Likert scale for several of the display's features and functions. A copy of the questionnaire is reproduced in the appendix. In addition, eye-tracking data were recorded during the experiment for later analysis.

The NASA TLX subjective workload questionnaire (Hart & Staveland, 1988) was administered to each pilot to obtain a subjective measure of workload for several of the in-flight tasks. NASA TLX is a six-item questionnaire measuring mental effort, physical effort, temporal pressure, perceived performance, total mental and physical effort, and frustration. Immediately after a flight, the pilot was asked to give a score (1 meaning low and 10 meaning high) for each of the six items on each of several tasks.

Flight Simulator

All of the flights were conducted in the Civil Aerospace Medical Institute's Basic General Aviation Research Simulator (BGARS), comprised of seven networked PCs configured using Microsoft's Flight Simulator 2002, with five out-the-window views and moderate-fidelity controls (Precision Flight Systems – Dual Professional Flight Console). Flight performance data were recorded four times per second and consisted of latitude, longitude, altitude, ground speed, heading, bank, cross-track error, elapsed time, and flight segment. The simulator was configured as a Cessna 182RG for this study. Figure 2 shows a picture of the simulator. Figure 3 shows how the GPS and MFD were situated in relation to the rest of the aircraft instruments.

Eye-tracker

The eye-tracking data were collected using the El-mar Vision 2000 Video Eye-tracking System. Horizontal and vertical eye position and pupil size were estimated at the rate of 60 Hz. An estimate of pupil center was used and compared with the positions of two corneal reflections to determine gaze location. Eye movements up to ±45° horizontal and up to ±35° vertical were measured by the system. The system was designed to compensate for head-band slippage of up to 5mm without affecting the accuracy of the eye position data. Data from the eye-tracker were recorded directly onto high-fidelity stereo video cassette recorder (VCR) tapes. These data included video from the scene camera and cursor, as well as encoded digital data comprising horizontal and vertical eye position, pupil size, blink, system status information, and frame count. The amount of time spent looking at four areas-of-interest (Instruments, Out-the-window, GPS/MFD, and Sectional

Figure 2. Basic General Aviation Research Simulator (BGARS).

Figure 3. Close-up view of the BGARS aircraft instruments and GPS/MFD displays (to the right).

Chart) was recorded using a frame-by-frame analysis of the video. These times were then converted to a percentage of time across the flight.

Procedure

Pilots were briefed about the experimental protocol and filled out consent forms. Each pilot received approximately 30 minutes of training on the GX-50 and the MX-20. Training on the GX-50 GPS display consisted primarily of recognizing and understanding the navigation information presented. Pilots were not required to interact with the GPS display controls during the flights, with the exception of responding to the appearance of messages.

Next, the pilots were shown how to use the MX-20 MFD. The focus of this training was on how to understand and interact with the custom moving-map display. Interaction with the display consisted of selecting one of the available map modes (i.e, track-up, track-up 360, north-up, or desired track-up) and selecting which types of overlay information to present. Pilots were also shown how to read and interpret the traffic information presented on the map display and how to call up and interact with the traffic page.

Following training, a 15-minute warm-up flight was given to the pilots to allow time for familiarization with the simulator, to ensure they could recognize traffic out-the-window, and to be sure they could perform the necessary flight tasks. At the conclusion of the warm-up flight, a break was offered to each pilot.

The experimental procedure required that the pilots fly two similar VFR flight plans in visual meteorological conditions. The flight plan for both flights consisted of taking off from the intial airport, flying directly over the top of a second airport, then flying straight toward the destination airport and landing. Before take-off, pilots were shown the route on the sectional chart and allowed to take the sectional chart with them during the flight. During scenarios where only conventional instruments were used, pilots would determine their headings and times using the sectional chart before they began the flight. During flights using the MFD and/or GPS, pilots were free, both before and during the flight, to set up and adjust those displays at their discretion.

The first flight scenario consisted of flying from the Manchester airport in Manchester, New Hampshire (MHT), to Marlboro airport (9B1) in Marlboro, Massachusetts, then direct to Norwood Memorial airport

(OWD) near Westwood, Massachusetts. Pilots were instructed not to land at the intermediate airport but simply to fly directly over the top of the airport before turning to the heading that would direct them to the destination airport. Figure 4 shows the area and route of the flight. The thick shaded lines show the basic flight segments used to measure flight path error during the flight.

The second flight scenario required the pilots to take off from Worcester Regional airport (ORH), fly over the top of the Fitchburg airport (FIT), and then fly directly to Beverly airport (BVY). Figure 5 depicts the flight area and route.

During both flights, pilots were instructed to circle around after taking off and fly directly over the runway before proceeding to the first waypoint. Before beginning each of the flights, pilots were given a list of 5 landmarks to identify during the flight. These landmarks consisted of airports, lakes, and a downtown area. Pilots were instructed to identify each landmark and then to provide an estimate of their current distance from the landmark. The weather was set to unlimited visibility for both flights. Winds were 10 knots from 260 degrees. In addition to the landmarks, pilots were also instructed to watch for and identify aircraft traffic during the flight. Each flight scenario lasted approximately 25 minutes. At the conclusion of the second flight, the usability questionnaire was administered to the pilots, and then they were debriefed.

Hypotheses

- Measures of workload would generally favor the advanced flight displays. This expectation was based on the presence of easily followed navigation information available on both the GPS display and MFD.
- Likewise, flight performance measures would generally favor the advanced flight displays. Prior research has shown that the use of horizontal situation displays such as the MX-20 improves navigation performance (Haskell & Wickens, 1993; Wickens & Prevett, 1995; Williams & Ball, 2003).
- Because the pilots would have a better awareness of their position during the flight, using the moving-map display would ease the task of identifying landmarks and estimating their current distance from those landmarks. Pilots using only the GPS might still benefit because they would be able to devote more of their

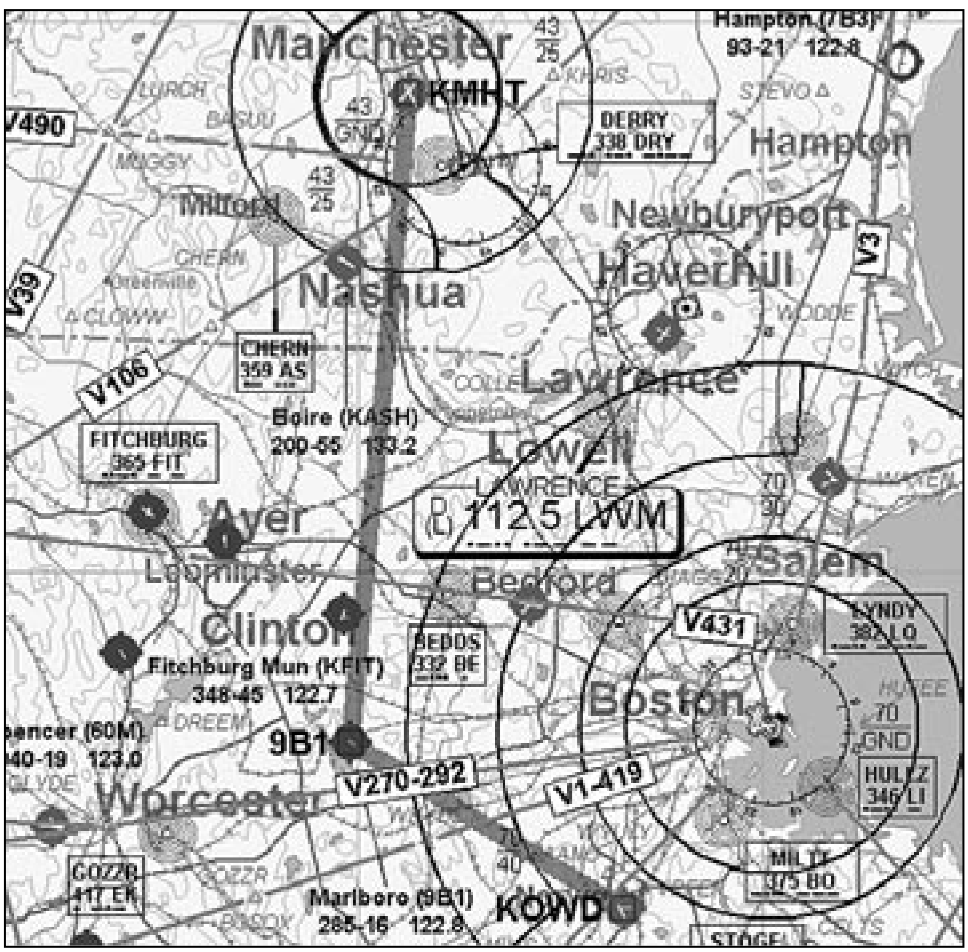

Figure 4. First flight scenario.

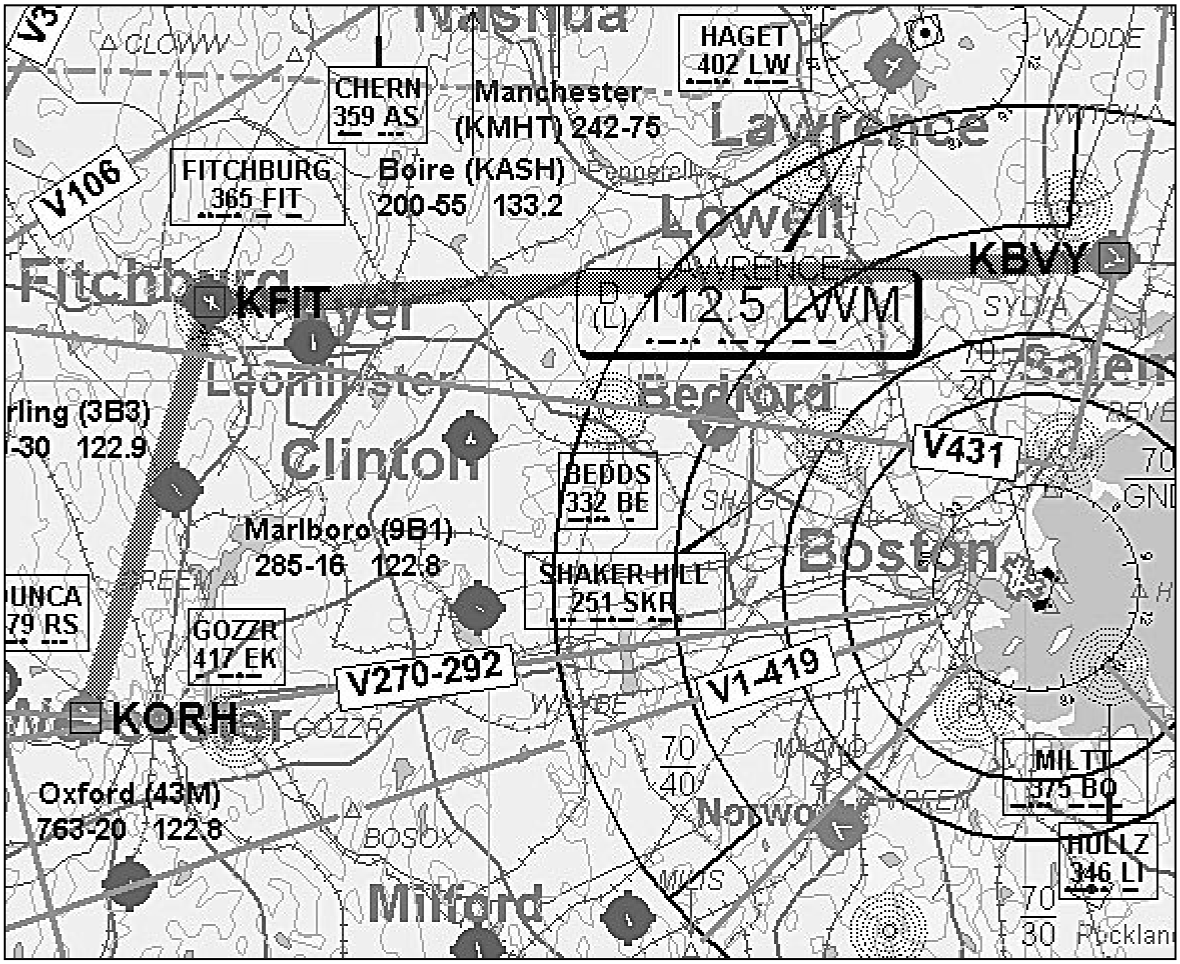

Figure 5. Second flight scenario.

mental resources to landmark identification and less to navigation than when flying with only conventional instrumentation.

- Finally, use of the traffic display would have beneficial effects on spotting traffic, but might have negative effects on scanning. The negative effect on scanning would be exacerbated by the general negative effect of advanced cockpit displays on scanning. This assumption was based on reports from pilots that fly these types of displays in real-world conditions (Williams, Yost, Holland, & Tyler, 2002).

RESULTS

Pilot Demographics

Flight hour estimates were collected for a number of flight categories. The mean, median, and range of each category are presented in Table 1.

In addition to flight hours, participants were asked to report on the types of aircraft they had flown. Three of the pilots (12.5%) had flown the type of aircraft modeled on the flight simulator (a Cessna 182RG). Eleven of the pilots (45.8%) had flown a complex aircraft with retractable gear and a variable pitch propeller. Participants were also asked whether they had experience with a GPS display. Seventeen of the pilots (70.8%) had used a GPS display. Mean hours use was 71.3. However, the sample was highly skewed, with one pilot reporting almost 1,400 hours of GPS display use. Median GPS use was 7 hours. Participants were asked whether they had used the MX-20 moving map display or a similar display. Two of the pilots (8.3%) had previously used the MX-20 display, and 11 pilots (45.8%) had used a moving map display similar to the MX-20.

Display Usability

All participants were asked to rate how easy it was to learn to operate the display system. Ratings were obtained on a 5-point scale from 1 (very difficult to learn) to 5 (very easy to learn). The mean rating was 4.17, with no rating below a 3 (i.e., average difficulty to learn). Finally, pilots were asked whether they would like to use the navigation system in a real aircraft, and all said they would like to use the system in real flight.

Table 1. Pilot Flight Time Demographics.

Demographic Category	Mean Hrs.	Median Hrs.	Range Hrs.
Total Time	840.46	502.5	135 – 3,850
Time: Last 12 Months	316.50	200	50 - 800
Time: Last 90 Days	101.09	80	0 - 250
Total Time: VFR	700.23	450	105 – 2,400
VFR Time: Last 12 Months	279.70	150	0 - 750
VFR Time: Last 90 Days	85.32	70	0 - 250
Total Time: Actual Instrument Flight	83.02	8	0 – 1,500
Actual Instrument Time: Last 12 Months	10.54	5	0 - 50
Actual Instrument Time: Last 90 Days	3.43	1	0 - 20
Total Time: Simulated Instruments	56.79	45	6 – 150
Simulated Instrument Time: Last 12 Months	26.33	22	0 - 100
Simulated Instrument Time: Last 90 Days	10.13	5	0 - 40

Table 2. MX-20 mean feature ratings on four usability criteria.

	Feature Rating Criteria				
Part 1. Multifunction Display	**Effect-iveness**	**Ease of Use**	**Cues & Prompts**	**Feed-back**	**Row Mean**
Accessing database information (runway lengths, frequencies, etc)	4.62	4.17	4.27	4.08	**4.29**
Map orientation selection (North up, Track up)	4.50	4.21	4.14	4.21	**4.27**
Selection of map data for display – decluttering	4.20	3.93	4.15	3.93	**4.05**
Using navigation data	4.29	4.20	4.36	4.29	**4.29**
Using the map range (zoom) feature	4.87	4.64	4.62	4.71	**4.71**
Using map information	4.40	4.00	4.15	4.07	**4.16**
Accessing traffic information page	3.92	3.67	4.00	4.08	**3.92**
Using traffic information	4.36	4.07	4.00	4.31	**4.19**
Selecting traffic display options	4.00	3.92	3.85	4.00	**3.94**
Column Mean	**4.35**	**4.09**	**4.17**	**4.19**	

After completing both flights, participants filled out the usability questionnaire. MX-20 and GX-50 features were rated according to each of 4 criteria along a 5-point scale. The criteria were effectiveness of the feature, ease of use, strength of cues and prompts, and effectiveness of feedback associated with the feature. A higher number indicated better usability for each of the criteria. The reader is directed to the Appendix for the definitions of the rating scale values. Table 2 presents the mean ratings for the MX-20 features on each of the 4 usability criteria.

In the table, the shading indicates the highest rating for a particular criterion. All features were rated very highly across all rating criteria. The table indicates that pilots believed that the map range (zoom) feature was the most usable aspect of the MX-20. During normal operation, two buttons were continuously dedicated for zooming, making it easy for the pilots to alter the zoom level. The lowest-rated features were concerned with accessing the traffic information page and selecting traffic display options. Both of these features were available only on a separate page from the custom map. The custom map was the initial page shown to pilots and was the page selected at the beginning of each flight. The majority of pilots (23 of 24, 96%) never selected the traffic page or traffic display options during the flight scenarios. For these pilots, their only exposure to the traffic page and traffic display options was during initial training.

In addition to the functions listed in Table 2, pilots were asked specific questions about the usability of a number of other features of the MX-20. Ratings ranged from 1 to 5 on a disagree-agree scale. These questions and their mean response values are listed in Table 3.

Looking at Table 3, we find that most of the items were rated highly by the pilots. Even the lowest-rated items were still rated neutral to positive. The two highest rated items were items 9 and 10, dealing with the ease of understanding button labels and the meaningfulness of colors on the moving-map display. The lowest-rated item was the statement, "Text displayed on the MFD was easy to read." Regarding this item, 5 of 16 pilots (31%, note: 8 of the 24 subject pilots did not use the MFD and did not rate items associated with the MFD) disagreed to some extent with the statement and only 63% agreed. One pilot wrote that text associated with the traffic was difficult to read. The second lowest-rated item, but also the item with the lowest percent agreement, was item 5, "Prompts for pilot inputs, alerts, and advisories effectively attracted my attention." It should be noted that pilot alerts and advisories were visual only, with no aural cue accompanying the alert. Of the alerts and advisories available on the MX-20, the only ones that the pilots regularly experienced during the experiment were a terrain warning flag, a traffic flag, and a barometric pressure confirmation.

Pilots also rated the GX-50 GPS features according to the same 4 usability criteria. Table 4 presents the mean feature ratings for the GX-50.

The GX-50 features, as with the MX-20, were rated very positively. As noted earlier, pilots did not interact with most of the functions of the GX-50.

The usability of additional features was addressed with specific questions. As before, these questions were rated on a disagree-agree scale of 1 to 5.

The GPS usability item receiving the lowest rating concerned the ability of the unit to attract pilot attention for warnings and alerts. The alerts received by pilots during the study were messages regarding the completion of flight segments and warnings regarding approach to restricted flight areas. As with the MX-20, no aural alerts were issued by the GPS.

Table 3. Mean scores and percent agreement for additional MX-20 (MFD) usability items.

1. The MFD information I needed was easy to see and not obscured by other information.	3.94	88%
2. Text displayed on the MFD was easy to read.	3.56	63%
2. The meaning of symbols on the MFD was easy to understand.	4.13	88%
3. The meaning of text on the MFD was easy to understand.	4.00	75%
4. Prompts for pilot inputs, alerts, and advisories effectively attracted my attention.	3.63	56%
5. Prompts for pilot inputs, alerts, and advisories were easy to understand.	4.13	81%
6. The MFD menus were easy to use.	4.06	81%
7. The MFD menu options were easy to understand.	4.13	88%
8. The MFD button labels were easy to understand.	4.25	88%
9. The use of colors on the MFD maps was meaningful.	4.31	81%
10. The size of the MFD display area was adequate for the information presented.	4.00	75%

Table 4. GX-50 mean feature ratings on four usability criteria.

	Feature Rating Criteria				
Part 2. GPS Navigator	**Effect-iveness**	**Ease of Use**	**Cues & Prompts**	**Feed-back**	**Row Mean**
Using the CDI data	4.15	4.00	4.10	4.20	**4.11**
Using cross-track error (XTE) data	3.89	4.00	3.95	4.00	**3.96**
Column Mean	**4.02**	**4.00**	**4.02**	**4.10**	

Table 5. Mean scores and percent agreement for additional GX-50 (GPS) usability items.

13. The knobs and buttons on the GPS receiver were placed in logical and convenient locations.	4.13	79%
14. The GPS receiver controls for different functions operate in a consistent manner.	4.04	75%
15. Text on the GPS display was easy to read.	4.50	92%
16. The size of the GPS display area was adequate for the information displayed.	4.25	79%
17. GPS warnings and alerts were issued at the right time.	4.21	79%
18. GPS warnings and alerts effectively attracted my attention.	3.75	67%

In addition to the specific usability questions, participants were asked if they had made any errors while interacting with the MX-20 and GX-50 displays. Six (37.5%) of the participants stated they had made errors with the MX-20. Seven (29.2%) participants admitted making errors with the GX-50. Mistakes with the MX-20 included confusions caused by flying south while using the display in a north-up orientation (2 pilots), forgetting to use the display while estimating distances (1 pilot), pushing incorrect buttons while interacting with the display (2 pilots), and misinterpreting information on the display (1 pilot). Mistakes with the GPS included difficulty with noticing and reacting to the message notification (3 pilots), trouble interpreting the information presented (3 pilots), and problems with pressing the wrong button (1 pilot).

Navigation Performance

The effect of the various displays on navigation performance was measured using both subjective workload estimates and objective performance measures. At the conclusion of each flight, pilots were asked to estimate the workload related to navigating during the flight. The 6 individual TLX workload estimates were averaged into a single composite workload estimate. Figure 6 presents a bar graph of the composite workload scores for navigating for each of the 3 types of display comparisons.

Paired t-tests were used to compare workload scores within each display comparison group. Results showed that pilots felt that workload associated with navigating the aircraft was significantly greater using conventional instruments (Con) than using the MX-20 multifunction display (MFD), $t(7) = 3.86$, $p < .05$. Likewise, pilots felt that workload using conventional instruments was significantly greater than using the GPS display, $t(7) = 3.42$, $p < .05$. On the other hand, the difference in estimated workload between using the GPS display and the GPS/MFD combination was not significant, $t(7) = 2.00$, $p = .09$.

To support the subjective workload estimates, pilot navigational performance was measured using the root mean squared error (RMSE) of the horizontal distance of the aircraft from the intended flight path and the RMSE of the altitude difference from the intended flight altitude. Horizontal error was measured for the 2 straight-line segments shown in Figures 4 and 5, and altitude error was measured for those 2 line segments and for the turn connecting the segments. Figure 7 shows the horizontal RMSE for each of the display comparison conditions.

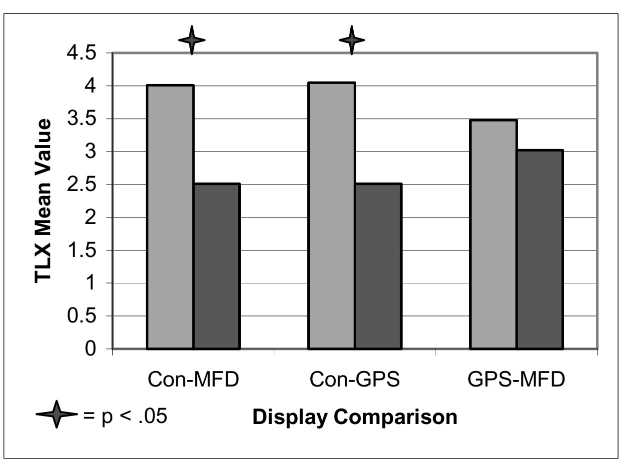

Figure 6. Composite NASA TLX subjective workload estimates for navigation by display comparison group.

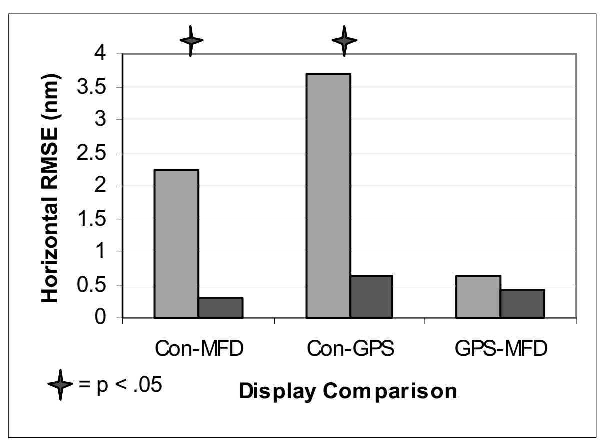

Figure 7. Horizontal navigation error, in nautical miles, across display comparison conditions.

An analysis of the horizontal RMSE measures showed a correspondence with the pilot workload estimates for navigation. Horizontal error was significantly greater when flying conventional instruments than when using the GPS/MFD combination, $t(7) = 3.05$, $p < .05$, or when using the GPS alone, $t(7) = 2.61$, $p < .05$. However, the difference in horizontal error between using the GPS and the GPS/MFD combination was not significant, $t(7) = .77$, $p = .46$.

An analysis of the altitude error did not show the same pattern of results. For all three display comparison groups, there were no significant differences found in altitude RMSE; $t(7) = .65$, $p = .53$ (Con-MFD); $t(7) = 1.83$, $p = .11$ (Con-GPS); and $t(7) = 1.51$, $p = .18$ (GPS-MFD). It should be noted here that, although the MFD provided a digital altitude display for the pilot, neither the MFD nor the GPS provide any type of vertical situation display.

Landmark Identification

The effect of the displays on the ability to locate and identify landmarks out-the-window was measured using both subjective workload estimates and objective measures. Figure 8 shows the mean TLX composite values for each of the three display comparison groups. Only for the Con-GPS group was there a significant difference in the estimated workload associated with the identification of landmarks, $t(7) = 3.03$, $p < .05$. Both the Con-MFD and GPS-MFD groups did not differ significantly in their workload estimates, $t(7) = 1.02$, $p = .34$ (Con-MFD) and $t(7) = 2.06$, $p = .08$ (GPS-MFD).

The objective measure for landmark identification was the number of landmarks correctly identified. Figure 9 shows the mean number of landmarks identified for each display comparison group.

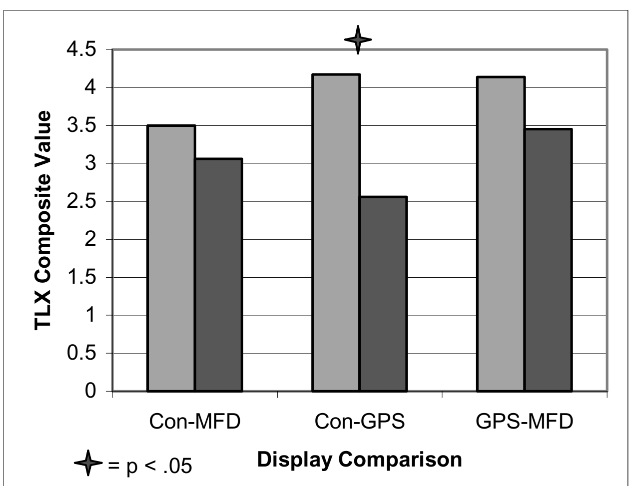

Figure 8. Composite NASA TLX subjective workload estimates for identifying landmarks, across display comparison conditions.

The pattern of results is consistent with the subjective workload estimates. The only display comparison condition where there was a significant difference in the number of landmarks identified was the Con-GPS condition, $t(7) = 5.02$, $p < .05$. There was not a significant difference in the number of landmarks identified for the Con-MFD, $t(7) = 1.36$, $p = .22$, or GPS-MFD conditions, $t(7) = .84$, $p = .43$.

Estimating Distance to Landmarks

Figure 10 shows the mean subjective workload scores for estimating the distance from a landmark. The pattern of workload estimates mirrors almost exactly the workload estimates for identifying landmarks. That is, the only group with a significant difference in workload estimates was the Con-GPS group, $t(7) = 4.40$, $p < .05$.

However, the workload estimates of the other two groups did not differ significantly, $t(7) = .74$, $p = .48$ (Con-MFD) and $t(7) = 1.86$, $p = .11$ (GPS-MFD). This pattern of results matches both the subjective workload estimates for locating landmarks and the success in locating landmarks; however, as we will see below, it does not match actual performance in the estimation of landmark distances.

An analysis of the RMSE scores of the estimated landmark distances found that none of the display comparison conditions showed a significant difference in the estimated-distance-to-landmark error. The mean across all conditions was between 2 and slightly more than 3 nautical miles. Paired t-test results were $t(7) = .59$, $p = .57$ (Con-MFD), $t(7) = .83$, $p = .43$ (Con-GPS) and $t(7) = .51$, $p = .63$.

It was evident during the experiment that pilots did not rely on either the GPS or the MFD to assist them in their distance estimates. For example, one of the landmarks was a lake located right beside the Marlboro airport, which was the first waypoint of the first flight scenario. Since the GPS was programmed with the location of the Marlboro airport, there was a continuous indication of the distance to that airport during the first leg of the flight that appeared on both the GPS display and the moving-map display. However, after sighting the lake, pilots did not use the distance indication on the GPS display or on the MFD to estimate the distance to the lake, which was within ½ mile of the airport.

Identifying Traffic

Pilots were asked to watch for and identify traffic during their flights. Following completion of the scenarios, pilots were asked to rate the subjective workload associated with locating traffic out-the-window. When analyzing the data, only the display comparison conditions containing the traffic display were included.

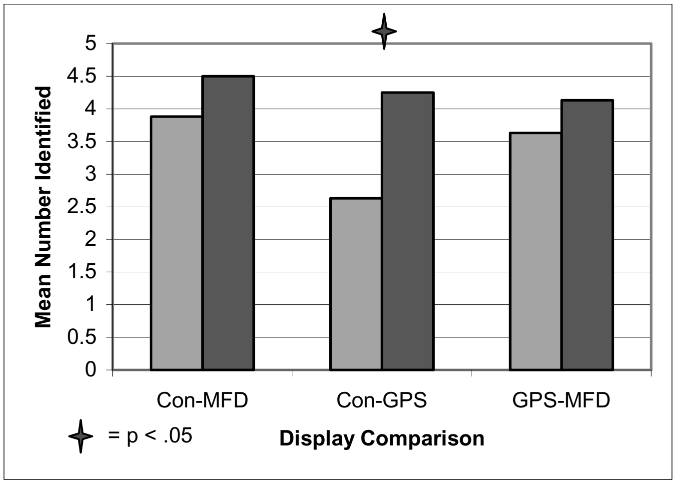

Figure 9. Mean number of landmarks identified across display comparison conditions.

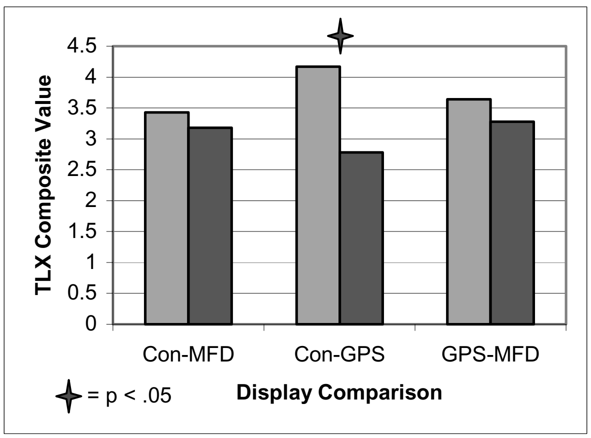

Figure 10. Composite NASA TLX subjective workload estimates for estimating landmark distance across display comparison conditions.

There were no significant differences in the workload estimates across display conditions, $t(7) = 1.07$, $p = .32$ (Con-MFD) and $t(7) = .41$, $p = .70$ (GPS-MFD). Pilots were not provided with any feedback regarding their success at spotting traffic; however, pilots with the traffic display could quite often see traffic on the display that would not be visible out-the-window. This could account for the slightly higher workload estimates found in the MFD conditions.

Actual success for locating traffic was computed by counting the number of times that an aircraft was correctly located in the out-the-window view (hits) and subtracting the number of times that an aircraft appeared in the out-the-window view but was not spotted (misses). Figure 11 shows the mean traffic locating score across display comparison conditions.

Of the two comparison conditions that included the traffic display (MFD), only one, the GPS-MFD condition, showed a significant positive effect for locating traffic when using the traffic display $t(7) = 2.28$, $p < .05$. There was not a significant improvement in locating traffic under the Con-MFD condition $t(7) = .31$, $p = .76$.

In addition to the ability to locate traffic, there was a question regarding the effect that the traffic display had on pilot scanning behavior, particularly as it related to looking out the window. An analysis of the eye-scan data was performed for the display comparison conditions in which the traffic display was used and the percentage of time that pilots looked out-the-window was computed. Figure 12 shows the mean percentage that pilots in each display condition directed their scan outside of the cockpit.

Pilots in the Con-MFD condition spent significantly more time scanning outside while flying conventional instruments, $t(5) = 3.05$, $p < .05$. There was no significant difference in the amount of time pilots spent looking out-the-window in the GPS-MFD comparison condition,

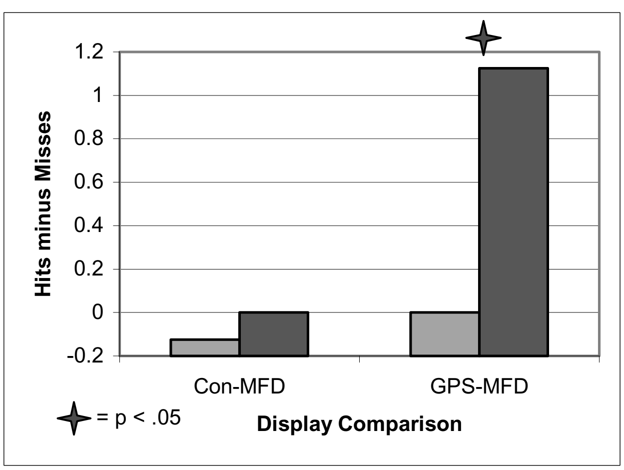

Figure 11. Mean traffic locating score (hits minus misses) across display comparison conditions.

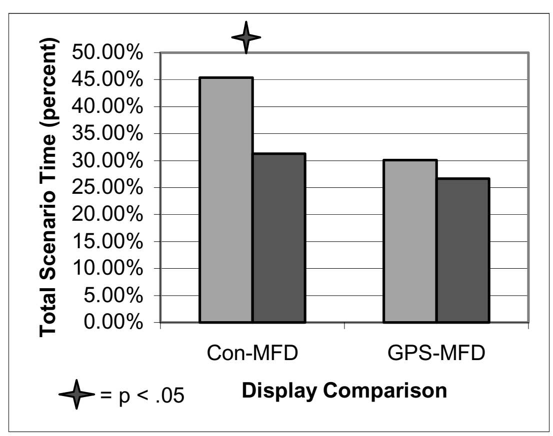

Figure 12. Percentage of time pilots spent looking out-the-window, across display comparison conditions.

$t(6) = 1.31$, $p = .24$. This was likely due to pilots using both the GPS and MFD to a greater extent relative to conventional instruments. Note that some of the subject data could not be analyzed for both of these comparison conditions because of eye-tracker malfunctions.

What is apparent from the eye-tracking data is that scanning behavior does not appear to be directly related to traffic detection performance. Pilots with conventional instruments looked outside significantly more while flying without the traffic display, but no difference in detection performance was found. On the other hand, pilots in the GPS-MFD condition spent relatively the same amount of time looking out-the-window whether using the traffic display or not, but a significant improvement in traffic detection was obtained while using the traffic display.

DISCUSSION

Results from the usability questionnaire demonstrated that pilots found both the GX-50 and MX-20 units to be usable, without any serious usability problems. This result matches earlier research on these units within an IFR environment (Williams & Ball, 2003). As with the first experiment, lower ratings were associated with the warning and alert aspect of the displays. The inclusion of an aural warning and alert that accompanied the visual flags would greatly improve the salience of these features. One user interface aspect that was rated more poorly in this experiment was the readability of the text that appeared on the MX-20 moving-map display. Pilots in the first experiment did not use the moving-map to assist in the identification of landmarks and therefore were not trying to read items on the display as they were in the present experiment.

As expected, subjective workload measures related to navigation and identifying landmarks generally favored the advanced flight displays. These results corresponded with the objective measures of horizontal navigation performance but not with the vertical measures of performance. This finding confirms other research showing that the presence of a moving-map display makes the navigation task easier by integrating information that was previously separated (Wickens & Carswell, 1995; Wickens, Liang, Prevett, & Olmos, 1996).

The workload estimates generally mirrored pilot performance for both navigation and locating landmarks out-the-window. The workload estimates for estimating landmark distance were similar to performance in locating landmarks but did not reflect how well they estimated distance. A likely explanation for this is that, because they did not receive feedback regarding their distance estimation performance, they based their workload estimates on their perceived performance regarding locating landmarks, which was a more salient task that did not require as much explicit feedback.

One unexpected finding was that subjective workload related to locating traffic out-the-window was not reduced by the presence of a traffic display. One possible explanation for this was the lack of feedback provided to the pilots regarding their success or failure at finding traffic. Most pilots saw only one or two aircraft during a scenario, and pilot comments following the flights confirmed that they expected to see more. Pilots were not told whether or not they had actually failed to see traffic appearing on the outside views during the scenario. This perceived lack of success was exacerbated by the presence of the traffic display showing several aircraft in close proximity during the scenarios.

As expected, the presence of an advanced navigation display resulted in superior navigation performance. This was true whether the pilots were using the MX-20 moving-map display or just the GX-50 GPS display. The advanced navigation displays also resulted in better performance in locating landmarks; however, the results here were mixed. Pilots flying the Con-GPS condition had better performance in locating landmarks while using the GPS. This was primarily due to poorer navigation performance while flying with conventional instruments, relative to the other display conditions. Pilots flying conventional instruments in the Con-MFD condition did not have as much navigational error as pilots in the Con-GPS condition. This probably led to the lack of a difference in locating landmarks for the Con-MFD condition. The ability to locate landmarks was the same whether using the MFD moving-map or only the GPS display.

Despite improving the ability to locate landmarks, the presence of an advanced navigation display did not improve the ability to estimate landmark distance. In fact, pilots rarely attempted to use the navigation displays in making their distance estimates. Perhaps some specific training would have changed this behavior.

The presence of a traffic display improved the ability to locate traffic out-the-window but not for all of the display comparison conditions. There was a confounding factor present in the experiment that prevented an unambiguous measurement of the effect of the traffic display on the ability to locate traffic. Traffic routing was matched to the intended flight path of the aircraft. Therefore, when pilots erred regarding their navigation, the amount of traffic that was encountered was altered. Unfortunately, both navigational accuracy and traffic locating ability could not be measured reliably within the experimental paradigm.

Even though the presence of the traffic display improved the ability to locate traffic out-the-window, eye-gaze data indicated that pilots spent less time looking out the window when using the traffic display. This finding suggests that a traffic display can make the pilot's scan more efficient. However, there is a potential danger with the display in cases where not all traffic that is present outside appears on the display. This condition is called mixed-equipage and can occur because these types of displays require an onboard piece of equipment called a Universal Access Transceiver (UAT) for the aircraft to be able to broadcast its position. Pilots flying in the Bethel, Alaska, area reported that there were times when they were surprised by an aircraft that did not appear on their display because that aircraft did not have the proper equipment for broadcasting its own position (Williams et al., 2002).

As these new aircraft displays become more common in general aviation cockpits, it is comforting to find evidence of their effectiveness and usability in both instrument and visual flight conditions. Issues still remain, however. One pressing question is the effect that mixed equipage will have on scanning and the ability to find aircraft out-the-window. More important is the effect of mixed equipage on a pilot's mental representation of traffic in the operating area and how that influences the allocation of the pilot's attention. Future research should address these questions. Other questions of importance are the type of traffic conflict alert that should be presented to pilots, how conflict resolution occurs, and how these types of traffic displays will impact aircraft operating with the traffic conflict avoidance system (TCAS). In addition, there is a question regarding the effect that losing a moving-map display would have on the ability to maintain position awareness and to complete a flight. The larger navigational errors that occurred when pilots flew using conventional instruments provides evidence of the difficulty with navigation that would be experienced if pilots lost their advanced navigation displays. If pilots rely too heavily on the moving-map display, and do not maintain a backup procedure, they could easily lose their position awareness if the display fails during the flight. A future study is needed to look at this issue.

REFERENCES

FAA (2000). US civil airmen statistics. Information available on the Internet at URL http://api.hq.faa.gov/CivilAir/index.htm. Web site of the FAA Office of Aviation Policy and Plans.

FAA (2004). FAA Alaska Region Home Page. Information available on the Internet at URL http://www.alaska.faa.gov/capstone/. Web site of the FAA Alaska Region Capstone Office.

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical, & Theoretical Research. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload,* North-Holland, Amsterdam, The Netherlands: Elsevier Science Pub.

Haskell, I. D. & Wickens, C.D. (1993). Two- and three-dimensional displays for aviation: A theoretical and empirical comparison. *International Journal of Aviation Psychology*, *3*(2), 87-109.

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

Wickens, C.D. & Carswell, C.M. (1995). The proximity compatibility principle: Its psychological foundation and its relevance to display design. *Human Factors*, *37*, 473-94.

Wickens, C.D., Liang, C.C., Prevett, T., & Olmos, O. (1996). Electronic maps for terminal area navigation: Effects of frame of reference and dimensionality. *International Journal of Aviation Psychology*, *6*, 241-71.

Wickens, C.D. & Prevett, T.T. (1995). Exploring the dimensions of egocentricity in aircraft navigation displays. *Journal of Experimental Psychology: Applied*, *1*(2), 110-35.

Williams, K.W. (2002). The Alaska Capstone Program: Increasing safety through technology. *The Federal Air Surgeon's Medical Bulletin*, Spring, 2002. US Department of Transportation, Federal Aviation Administration, pp. 6-7.

Williams, K.W. & Ball, J.D. (2003). *Usability and effectiveness of advanced general aviation cockpit displays for instrument flight procedures*. U.S. Department of Transportation, Federal Aviation Administration, Office of Aerospace Medicine, Washington, DC. Publication No. DOT/FAA/AM-03/17.[1]

Williams, K.W., Yost, A., Holland, J., & Tyler, R.R. (2002). *Assessment of advanced cockpit displays for general aviation aircraft –The Capstone Program*. U.S. Department of Transportation, Federal Aviation Administration, Office of Aerospace Medicine, Washington, DC. Publication No. DOT/FAA/AM-02/21.[1]

[1]This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications Web site: http://www.cami.jccbi.gov/aam-400A/index.html

APPENDIX A
Usability Questionnaire

GPS/MFD Full-Mission Usability Study: Visual Flight Rules

1. Please rate the usability of each feature in single-pilot VFR operations using the rating scales in the boxes below. For example, for the first feature, choose one of the items numbered 1 – 5 under A, Easy or difficult to use, and write that number in Box A to the right of the first feature description.

A. Effectiveness of Feature/Function 1 = Very ineffective 2 = Moderately ineffective 3 = Borderline effective/ineffective 4 = Moderately effective 5 = Very effective	C. System cues and prompts 1 = Very ineffective 2 = Moderately ineffective 3 = Borderline effective/ineffective 4 = Moderately effective 5 = Very effective
B. Easy or difficult to use 1 = Very difficult 2 = Somewhat difficult 3 = Borderline 4 = Easy to use 5 = Very easy to use	D. System feedback following my actions 1 = Very ineffective 2 = Moderately ineffective 3 = Borderline effective/ineffective 4 = Moderately effective 5 = Very effective

	Feature Rating Criteria			
Part 1. Multifunction Display	**A**	**B**	**C**	**D**
Accessing database information (runway lengths, frequencies, etc)				
Map orientation selection (North up, Track up)				
Selection of map data for display – decluttering				
Using navigation data				
Using the map range (zoom) feature				
Using map information				
Accessing traffic information page				
Using traffic information				
Selecting traffic display options				

Please feel free to write a few words below each of the following items to explain your rating.

1. The MFD information I needed was easy to see and not obscured by other information.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

2. Text displayed on the MFD was easy to read.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

3. The meaning of symbols on the MFD was easy to understand.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

Usability Questionnaire (Continued)

4. The meaning of text on the MFD was easy to understand.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

5. Prompts for pilot inputs, alerts, and advisories effectively attracted my attention.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

6. Prompts for pilot inputs, alerts, and advisories were easy to understand.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

7. The MFD menus were easy to use.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

8. The MFD menu options were easy to understand.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

9. The MFD button labels were easy to understand.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

10. The use of colors on the MFD maps was meaningful.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

11. The size of the MFD display area was adequate for the information presented.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

12. Did you make any errors in using the MFD (any misinterpretations of displayed information, or anything you did incorrectly or omitted)? Yesî No î
    If so, please describe the error(s) as well as the cause and consequences of those errors:

A. Effectiveness of Feature/Function	C. System cues and prompts
1 = Very ineffective 2 = Moderately ineffective 3 = Borderline effective/ineffective 4 = Moderately effective 5 = Very effective	1 = Very ineffective 2 = Moderately ineffective 3 = Borderline effective/ineffective 4 = Moderately effective 5 = Very effective
B. Easy or difficult to use 1 = Very difficult 2 = Somewhat difficult 3 = Borderline 4 = Easy to use 5 = Very easy to use	D. System feedback following my actions 1 = Very ineffective 2 = Moderately ineffective 3 = Borderline effective/ineffective 4 = Moderately effective 5 = Very effective

Part 2. GPS Navigator	**A**	**B**	**C**	**D**
Using the CDI data				
Using cross-track error (XTE) data				

Please feel free to write a few words below each of the following items to explain your rating.

13. The knobs and buttons on the GPS receiver were placed in logical and convenient locations.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

14. The GPS receiver controls for different functions operate in a consistent manner.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

15. Text on the GPS display was easy to read.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

16. The size of the GPS display area was adequate for the information displayed.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

17. GPS warnings and alerts were issued at the right time.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

18. GPS warnings and alerts effectively attracted my attention.

1	2	3	4	5
Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree

19. Did you make any errors in using the GPS (any misinterpretations of displayed information, or anything you did incorrectly or omitted)? Yes No
If so, please describe the error(s) as well as the cause and consequences of those errors:

Comparing flying the first scenario to flying the second scenario:

20. My ability to navigate during the second scenario was:

1	2	3	4	5	6	7
Much Worse	Worse	Somewhat Worse	Same	Somewhat Better	Better	Much Better

21. My ability to identify landmarks during the second scenario was:

1	2	3	4	5	6	7
Much Worse	Worse	Somewhat Worse	Same	Somewhat Better	Better	Much Better

22. My ability to estimate distance to landmarks during the second scenario was:

1	2	3	4	5	6	7
Much Worse	Worse	Somewhat Worse	Same	Somewhat Better	Better	Much Better

23. My ability to remain aware of my ground speed, course, and altitude during the second scenario was:

1	2	3	4	5	6	7
Much Worse	Worse	Somewhat Worse	Same	Somewhat Better	Better	Much Better

24. My ability to locate traffic out the window during the second scenario was:

1	2	3	4	5	6	7
Much Worse	Worse	Somewhat Worse	Same	Somewhat Better	Better	Much Better