Portable Weather Applications for General Aviation Pilots

Ulf Ahlstrom, Federal Aviation Administration, Atlantic City International Airport, New Jersey, **Oliver Ohneiser**, German Aerospace Center, Braunschweig, Germany, and **Eamon Caddigan**, TG O'Brien & Associates, Inc., Westcliffe, Colorado

Objective: The objective of this study was to examine the potential benefits and impact on pilot behavior from the use of portable weather applications.

Method: Seventy general aviation (GA) pilots participated in the study. Each pilot was randomly assigned to an experimental or a control group and flew a simulated single-engine GA aircraft, initially under visual meteorological conditions (VMC). The experimental group was equipped with a portable weather application during flight. We recorded measures for weather situation awareness (WSA), decision making, cognitive engagement, and distance from the aircraft to hazardous weather.

Results: We found positive effects from the use of the portable weather application, with an increased WSA for the experimental group, which resulted in credibly larger route deviations and credibly greater distances to hazardous weather (≥30 dBZ cells) compared with the control group. Nevertheless, both groups flew less than 20 statute miles from hazardous weather cells, thus failing to follow current weather-avoidance guidelines. We also found a credibly higher cognitive engagement (prefrontal oxygenation levels) for the experimental group, possibly reflecting increased flight planning and decision making on the part of the pilots.

Conclusion: Overall, the study outcome supports our hypothesis that portable weather displays can be used without degrading pilot performance on safety-related flight tasks, actions, and decisions as measured within the constraints of the present study. However, it also shows that an increased WSA does not automatically translate to enhanced flight behavior.

Application: The study outcome contributes to our knowledge of the effect of portable weather applications on pilot behavior and decision making.

Keywords: decision making, flight displays, mobile devices, navigation, situation awareness

---

Address correspondence to Ulf Ahlstrom, William J. Hughes Technical Center, Federal Aviation Administration, Atlantic City International Airport, NJ 08405, USA; e-mail: Ulf.Ahlstrom@FAA.gov.

HUMAN FACTORS
Vol. 58, No. 6, September 2016, pp. 864–885
DOI: 10.1177/0018720816641783

INTRODUCTION

Visual flight rules (VFR) flight into instrument meteorological conditions (IMC)—where pilots inadvertently enter clouds or haze and can no longer see the horizon or the terrain—is a major safety hazard for general aviation (GA) pilots (Goh & Wiegmann, 2001). This dangerous situation can lead to spatial disorientation whereby pilots lose control of the aircraft (Wiggins, Hunter, O'Hare, & Martinussen, 2012; Wilson & Sloan, 2003). Some of the underlying causal factors in VFR flight into IMC relate to pilot characteristics, like experience and risk tolerance (Wiggins et al., 2012). Other factors relate to decision making and the ability of pilots to detect, incorporate, and respond to cockpit and "out-the-window" information and understand the potential effect of forecasted weather conditions (O'Hare & Stenhouse, 2009; Wiegmann, Goh, & O'Hare, 2002; Wiggins, Azar, Hawken, Loveday, & Newman, 2014).

There are many weather conditions that are hazardous to GA aircraft. For example, thunderstorms can produce lightning, heavy rain, hail, turbulence, icing, and wind shear. Areas surrounding a storm can also yield obstructions to visibility and reduce a pilot's ability to perceive the layout of runways, surrounding terrain, or the position of other aircraft. The knowledge of these weather conditions is especially important as pilots must understand the impact of weather on flights. However, as reported by Dutcher and Doiron (2008), pilots in general do not perform well on weather tests. Pilots often lack operationally relevant weather skills and frequently overestimate their weather knowledge. Dutcher and Doiron conclude that a key contributor to the lack of operationally relevant weather skills is the minimal weather training, a mere 9 hr, provided to private pilots in U.S. ground schools.

To prepare for a flight and to avoid encounters with hazardous weather, GA pilots can get a general overview of weather conditions by reviewing weather information on various Internet or television weather sites. If conditions are favorable for flight, pilots can acquire a more in-depth weather briefing by contacting a Flight Service Station (FSS). At the FSS, specialists can provide pilots with a detailed weather briefing, available weather forecasts, and weather reports that describe weather conditions along the intended route of flight. In addition, pilots can receive weather information from automated FSSs (AFSSs), which provide continuous telephone recordings of meteorological information. Pilots can also use the Web- or phone-based Direct User Access Terminal (DUAT) service to get weather information and flight-plan processing services.

During VFR flights, pilots navigate and avoid hazardous weather by visual sampling from the out-the-window view. Pilots integrate this visual information with data from the aircraft instruments, information from weather-reporting facilities, and information from other pilots on the radio frequency. Pilots can also receive weather information and request "flight following" by contacting air traffic control (ATC). Because VFR flights are based on "see and avoid," pilots must maintain minimum VFR cloud clearances throughout the flight. During instrument flight rules (IFR) flights, when visibility prevents visual sampling from the out-the-window view, pilots navigate by means of cockpit instruments and follow directions from ATC. Pilots must avoid in-flight weather hazards by either using cockpit weather systems, by requesting weather information from ATC, or by receiving information from AFSSs.

One increasingly popular method for receiving in-flight weather updates is the use of cockpit-mounted weather displays, like certified installed display systems or commercially available weather displays. These displays allow GA pilots to receive important aircraft, terrain, and weather information while in flight (Zimmerman, 2013). Potentially, this weather information could help pilots maintain good situation awareness, enhance weather decision making, and reduce safety risks, like VFR flight into IMC (Wilson & Sloan, 2003).

Despite these technological improvements and the increasing popularity of GA weather displays, human-in-the-loop simulations sometimes fail to reveal a clear benefit and improved pilot weather decision making from the use of cockpit weather displays. For example, in studies in which pilots used Next Generation Radar (NEXRAD) displays, researchers found that pilots seemed to be affected by and changed their weather deviation behavior according to the spatial resolution of the precipitation display. Beringer and Ball (2004) investigated the use of NEXRAD displays with an 8-km, 4-km, or 2-km spatial resolution. Pilots who used the high-resolution NEXRAD displays attempted to navigate between weather more than pilots with low-resolution displays. Furthermore, 53% of the pilots failed to comply with the recommended 20-statute-miles-separation guidance from storms (Federal Aviation Administration [FAA] & National Oceanic and Atmospheric Administration [NOAA], 1983). This finding seems to imply that pilots' deviation behavior was caused by a difference in perceived affordances triggered by the precipitation graphics. A similar result has also been reported by Wu, Gooding, Shelley, Duong, and Johnson (2012) in a study of pilot decision making during convective weather. In most cases, pilots' closest point of approach to hazardous weather was well below 20 nautical miles [nmi], meaning that pilots failed to comply with the separation guidance.

Even more striking is the result from a GA pilot study by Burgess and Thomas (2004). They investigated the effect of improved cockpit weather displays on GA pilot decision making and weather avoidance. Two different displays were used: an improved weather display with NEXRAD image looping and an improved weather display with the National Convective Weather Forecast (NCWF) product. With the average minimum distance from the aircraft to hazardous precipitation cells (50+ dBZ) as one of the dependent measures, the results showed no meaningful difference in weather avoidance between a control group (no weather display) and two groups using either NEXRAD image looping or the NCWF product. The mean minimum distance to hazardous cells for all three pilot groups was roughly 10 nmi—only half of the recommended safety margin. In addition to research on

NEXRAD displays, research has also shown issues with pilots' interpretation of cockpit radar displays. Unlike NEXRAD displays, which present a mosaic from multiple radar sites on the ground, radar displays depict hazardous intensities as measured by airborne radars. Wiggins et al. (2014) examined experienced pilots' ratings of turbulence levels associated with simulated radar displays. As a general rule, the precipitation intensity increases with an increase in turbulence. The result showed a lack of reliability in pilots' turbulence assessments. According to Wiggins et al. (2014), the fact that experienced pilots differ in how they interpret weather radar displays indicates a need for more comprehensive training on how to interpret weather radar in addition to display design improvements.

Researchers have also assessed the effect of advanced synthetic vision displays on pilot weather avoidance behavior. A study by Johnson, Wiegmann, and Wickens (2006) revealed that synthetic vision displays with terrain, weather, and moving maps do not help pilots avoid penetrating IMC clouds more than pilots using standard cockpit instruments. In fact, pilots who used advanced displays were 6 times more likely to fly using VFR into IMC and penetrate clouds. The researchers attributed this effect to attentional tunneling, whereby pilots using advanced displays were less likely to sample weather information from the out-the-window view. As a possible remedy for the attentional tunneling effect, the researchers propose a more effective display design along with pilot scan pattern training.

Besides precipitation information, modern cockpit weather displays can also contain information about winds, lightning, echo tops, and aviation routine weather reports (METARs), to mention a few. METAR information is especially important as it can provide the pilot with visibility, ceiling, and flight category information. In their study of advanced weather displays, Johnson et al. (2006) assessed pilot use of METAR symbols. They found a modest effect of the displayed METAR information, with only 6% of the pilots using METAR information strategically. Similarly, Coyne, Baldwin, and Latorella (2005) found that color-coded METAR symbols tended to bias pilots' estimates of ceiling and visibility. If the METAR symbols indicated more favorable conditions than what pilots could sample from the out-the-window information, pilots' visibility and ceiling reports were positively biased (and vice versa). O'Hare and Waite (2012) investigated the effect of symbol augmentation and found that pilots recalled more information from METARs when the information was presented with both text and symbols. Aside from pilot bias and a low usage of METAR symbols, research showed that METAR colors and METAR symbol shapes affect pilot detection of weather state changes during flight.

Ahlstrom and Suss (2015) investigated the effect of weather symbology on pilot ability to detect METAR status changes during simulated flights. They found a strong change-blindness effect (Nikolic, Orr, & Sarter, 2004; Rensink, 2000), with pilots varying considerably in their overall detection of METAR symbol display changes during flight. The overall detection performance ranged from 25% to 62% depending on the specific METAR symbol and color. Pilots who did not detect METAR changes (signaling reduced ceiling and visibility) at or around the destination airport were more likely to continue their VFR flight toward the preplanned destination. On the other hand, pilots who detected the METAR changes were more likely to request weather updates from ATC, consider alternate destination airports, or request an IFR flight plan. Similarly, Ahlstrom (2015b) investigated the effects of variations in cockpit weather display symbols and colors on GA pilot behavior, decision making, and cognitive engagement during a weather avoidance flight. The result showed behavioral and perceptual asymmetries in response to variations in weather symbology, with credible group differences for distance to weather, weather display usage, and cognitive engagement. In agreement with previous weather display studies, the outcome also showed that pilots came much closer to hazardous weather than what is recommended by current guidelines (i.e., 20 statute miles).

Besides certified weather display systems or commercially available weather displays for the cockpit, there is now a plethora of low-cost alternatives that can provide pilots with in-flight weather information. These alternatives rely on portable weather receivers and a subscription to commercial weather products that can be viewed on various portable devices or on cell phones. Besides providing weather information, many of these products also allow users to superimpose a route or flight plan on the weather map. The products also have an elaborate menu structure whereby the user can select a large number of graphical and text-based weather elements, forecasts, and weather briefings. Additionally, with the use of Automatic Dependent Surveillance–Broadcast, pilots can receive in-flight weather information that covers the Flight Information Services–Broadcast basic products with graphical and text-based information from Airmen's Meteorological Information, Significant Meteorological Information, METAR, NEXRAD, Notice to Airmen, pilot report (PIREPs), special-use airspace, Terminal Area Forecast (TAF), and wind, temperature, and lightning information.

Although portable devices have shown promise as a tool for navigation (Ware & Arsenault, 2012), there are reports of human factors issues that need to be resolved. One factor relates to the display of maps on portable devices. Because portable devices are small, there is a potential for increased display clutter, which can negatively affect user attention to details and the perception of state changes (Uluca, Streefkerk, Sciacchitano, & McCrickard, 2008). Hipp, Schaub, Kargl, and Weber (2010) have addressed user interaction problems with route selection, route deviation, and the implementation of warning messages on portable automotive navigation devices. The size, warning message, and the navigation problems uncovered are all issues that apply to portable cockpit navigation and portable weather avoidance navigation. Another issue with portable weather displays is the lack of standardized training. As pointed out by Dutcher and Doiron (2008), great strides have been made developing technology to improve flight efficiency and flight safety but at the cost of insufficient training. Portable weather applications are certainly becoming an important piece of current weather and display technologies. However, the training on how to use those weather applications is left up to the user. As weather training for private pilots is minimal, there is an even greater unknown how much training, if any at all, that private pilots have on the use of portable weather applications during flight.

A review of previous research reveals several main issues with GA pilot use of cockpit weather information. First, pilots often do not use weather display information in an operationally appropriate manner. Second, pilot decision making and behavior is affected by weather display factors, like resolution, color, and symbology. Third, despite using available weather information, pilots do not seem to maintain a safe distance from hazardous weather. Currently, there is a lack of research on the effect of GA pilot behavior from the use of portable weather displays. Therefore, the goal of the present study is to assess the effect of portable weather presentations on GA pilot behavior and weather situation awareness (WSA) during weather avoidance flights. We hypothesize that using a portable weather application could improve pilot WSA and assist pilots in avoiding areas of hazardous weather.

METHOD

Participants

Seventy GA pilots volunteered for participation in the study. The participants came from a group of pilots with a large variation in flight hours and ratings. For the simulation, we randomly assigned each pilot to an experimental group or a control group. The descriptive characteristics for both groups in terms of age and flight experience are shown in Table 1.

In Table 2, we provide a group summary of the most common pilot ratings and pilot certificates. In most cases, each individual pilot had multiple ratings.

Out of the 70 pilots, there were only 16 who had previous experience with cockpit weather displays (eight in each group). Of these 16 pilots, only eight pilots (three pilots in the control group and five pilots in the experimental group) had experience with portable weather displays, like Foreflight, SiriusXM, and WingX Pro. Four of these pilots (two in each group) were proficient—that is, they were actively using portable weather displays when flying as private pilots. Thirty-one of the pilots had taken additional meteorology classes beyond the basic weather training provided at ground schools. There were only two pilots in the control group and one pilot in the experimental group with prior training on how to

TABLE 1: Descriptive Characteristics of Study Pilots

				Flight Hours Accrued					
		Age (years)		Total		Instrument		Instrument–Last 6 Months	
Group	*n*	Median	Range	Median	Range	Median	Range	Median	Range
Experimental	36	54	21–86	4,650	90–30,000	1,000	0–14,000	2	0–120
Control	34	64	21–87	11,000	75–35,000	850	0–25,800	5	0–200

TABLE 2: Pilot Ratings

	Ratings							
Group	Private	Commercial	ATP	SEL	MEL	Instrument	CFI	CFI-II
Experimental	9	12	18	21	18	24	11	14
Control	9	12	20	26	23	18	11	14

Note. Private = private pilot certificate, Commercial = commercial pilot license, ATP = airline transport pilot certificate, SEL = single engine land certificate, MEL = multi-engine rating, Instrument = instrument rating, CFI = certified flight instructor, and CFI-II = certified flight instructor instrument.

interpret weather displays. Taken together, the study participants came from a diverse pool of pilots who mostly had very little (e.g., only a few flights) or no prior experience and training on electronic weather displays.

Simulation Equipment

GA cockpit simulator. The simulation was performed using a GA cockpit simulator configured as a Mooney Bravo single-engine aircraft. The simulator had an enclosed fuselage and was equipped with a 180° out-the-window view (see Figure 1). We used Microsoft Flight Simulator 2004 to control the flight characteristics of the simulator, Project Magenta G1000 type GA glass cockpit software to display the aircraft's control scheme, and the Lockheed Martin Prepar3D software to generate the out-the-window view.

Portable weather application. The portable weather application was specifically designed to provide a platform for weather display research and to be similar to commercially available products. The weather application design was based on user surveys, interviews, and cognitive walkthroughs. An initial survey of a cross-section of pilots, dispatchers, ATC, and traffic managers identified commonly used/requested weather information. These information items were then rated by frequency of use and the perceived importance. Subsequently, the weather items were prioritized based on their technical feasibility and the availability of weather data. To receive feedback on the final weather application content, researchers used the application in scenario-based cognitive walkthroughs using pilots as subjects (FAA, 2014).

The portable weather application ran on an iPad Air 2 that was attached to the participants by a leg strap (see Figure 2).

The weather information was overlaid on a map with a green line representing the preplanned route and a red plus sign representing the aircraft position symbol (see Figure 3). All distances and directions between waypoints were known to the pilots prior to flight, allowing pilots to estimate distances from the aircraft to areas of weather. Furthermore, pilots could use the cockpit's distance-measuring equipment to estimate distances to weather by selecting very-high-frequency omnidirectional radio range waypoints in close proximity to areas of weather.

An application menu allowed pilots to select weather information layers and weather

Figure 1. The cockpit simulator.

information features. The weather information features included aviation routine weather reports (METARs), TAFs, and PIREPs. By selecting the weather layers, pilots could display graphical area depictions of flight categories (e.g., VFR and IFR), ceiling information, visibility, precipitation, icing probability, turbulence potential, wind, temperature, relative humidity, and satellite imagery information. Each image was adjustable by means of a display zoom.

Functional near-infrared (fNIR) system. In the present study, we used objective fNIR recordings as a measure of cognitive engagement. In general, when cortical neurons are activated, there is a local increase in oxygenated hemoglobin and a decrease in deoxygenated hemoglobin, indicating an increased brain metabolism. Previously, we have used fNIR recordings during air traffic control simulations (Ahlstrom, 2015a; Harrison et al., 2014) and cockpit simulations (Ahlstrom & Suss, 2015). The continuous-wave fNIR system is connected to a flexible forehead sensor pad that contains four light sources (peak wavelengths at 730 nm and 850 nm) and 10 detectors. This configuration generates a total of 16 measurement locations or voxels per wavelength. With two wavelengths and dark current recordings for each of the 16 voxels, the system generates a total of 48 measurements for each 2 Hz sampling period.

Simulation Task

The current study was part of a larger project to assess the minimal weather requirements for GA pilots. During the simulation, we used two weather scenarios to assess pilot behavior in response to reduced visibility and convective activity. Here, we report the outcome for one of the scenarios that was specifically designed to assess the use of a portable weather application during convective weather avoidance. For the outcome of the visibility simulation, see Ahlstrom, Caddigan, Schulz, Ohneiser, Bastholm, and Dworsky (2015).

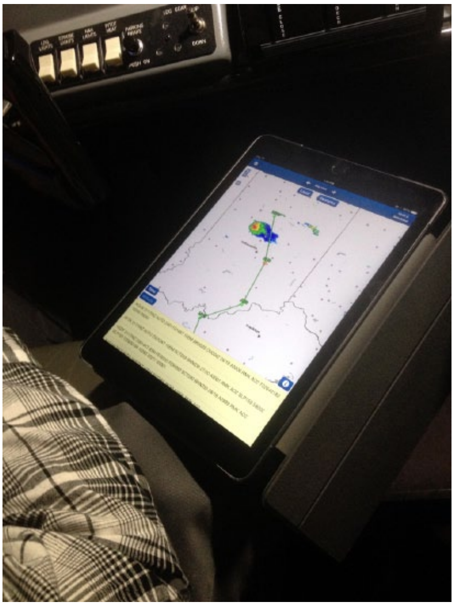
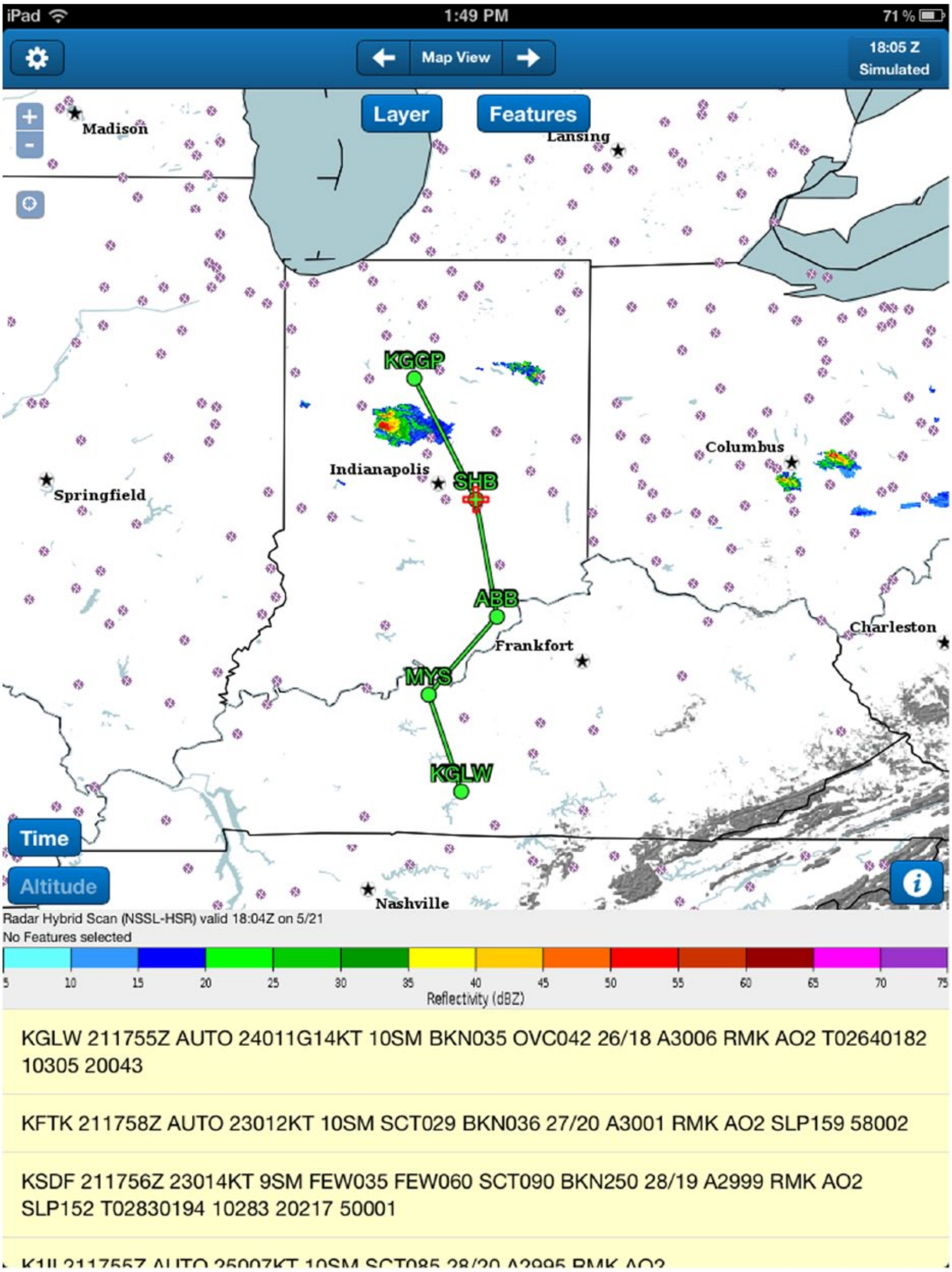


Figure 2. The portable weather display.

During the simulation, pilots flew a 20-min leg of a longer preplanned route starting at the airport at Glasgow Municipal, Glasgow, Kentucky (KGLW), and ending at the airport at Logansport/Cass County, Indiana (KGGP), as illustrated in Figure 3. During the scenario, pilots started midair (6,000 ft.) and encountered thunderstorms in the vicinity of KGGP. At the scenario startup, the weather conditions were visual meteorological conditions (VMC) with only some smaller cloud formations at altitude. As the flight progressed toward the destination, pilots encountered areas of marginal VMC. At lower altitudes near the destination airport, pilots had to avoid areas of IMC. An example of the out-the-window view during startup is illustrated in Figure 4.

Independent Variables

The only independent variable manipulated in this study was the availability of a portable



Radar Hybrid Scan (NSSL-HSR) valid 18:04Z on 5/21
No Features selected

KGLW 211755Z AUTO 24011G14KT 10SM BKN035 OVC042 26/18 A3006 RMK AO2 T02640182 10305 20043

KFTK 211758Z AUTO 23012KT 10SM SCT029 BKN036 27/20 A3001 RMK AO2 SLP159 58002

KSDF 211756Z 23014KT 9SM FEW035 FEW060 SCT090 BKN250 28/19 A2999 RMK AO2 SLP152 T02830194 10283 20217 50001

K1IL 211755Z AUTO 25007KT 10SM SCT085 28/20 A2995 RMK AO2

Figure 3. Illustration of the portable weather application with the route (in green), aircraft position symbol (red plus symbol), and precipitation information (yellow areas are ≥30 dBZ intensities).

Figure 4. Illustration of the out-the-window view at the scenario startup.

weather application. During the simulation, only the experimental group had access to this information.

Dependent Variables

During the simulation, we recorded six dependent measures, as outlined in Table 3.

In the following sections we describe how we calculated deviation and distance-to-weather measures. We also describe each dependent variable and clarify how it was measured during the simulation.

Computation of deviation and distance-to-weather measures. Current guidelines by FAA and NOAA (1983) state that hazardous weather should be avoided by at least 20 statute miles. In this study, we measured distance to weather (≥30 dBZ cells) to assess how pilot weather avoidance behavior is affected by the use of portable weather applications. We also measured the horizontal flight profile, defined as an aircraft's deviation from the preplanned route.

The flight scenario included convective activity with a southeast-moving weather cell, shown in the upper-left quadrant of the portable display map (refer to Figure 3). At 1-min intervals during the scenario, we measured the distance from the aircraft location (i.e., latitude/longitude) to the closest point of approach for ≥30 dBZ precipitation cells (visualized as yellow pixels). We also measured, once every 10 s, the aircraft position in relation to the preplanned route. The outcome of these measurements was the distances between aircraft and weather cells and the distances between aircraft and the preplanned route.

We used several defined parameters for the analyses. Our aircraft log files contained, among other data parameters, the elapsed scenario time in seconds, latitude, longitude, altitude, and heading. In a first step, the evaluation algorithm extracted all coordinates for each time interval (2 Hz) and saved them in a vector. Second, the algorithm loaded the scenario data with all the defined route points along with their latitude/longitude values. Third, the algorithm computed all distances in nautical miles, d, between latitude/longitude points as a great circle distance using the spherical law of cosines:

$$d = \mathrm{acos}(\sin[\mathrm{latA}] \cdot \sin[\mathrm{latB}] + \cos[\mathrm{latA}] \cdot \cos[\mathrm{latB}] \cdot \cos[\mathrm{lonB} - \mathrm{lonA}]) \cdot 3440.065.$$

Weather avoidance. To avoid flying into clouds or haze, pilots had to adjust their altitude and/or deviate from their preplanned route. We captured the vertical flight profile by analyzing pilots' altitude changes (in feet) once a second and the horizontal flight profile by recording pilots' deviations (in nautical miles) from the preplanned route every 10 s.

Decision making. We recorded all instances when pilots announced a decision to turn around or to use an alternate airport.

Distance to hazardous weather. To assess how pilot weather avoidance behavior was affected by the use of portable weather applications, we measured the distance to weather (≥30 dBZ cells) once a minute. At the end of the scenario, we asked pilots to rate how easy it was to determine the location of severe precipitation areas (7-point scale; 1 = *very difficult*, 7 = *very easy*).

TABLE 3: Dependent Measures

Number	Dependent Variable	Description
1	Weather avoidance	The flight profile: Altitude Horizontal deviation
2	Decision making	Pilot decision to deviate from the preplanned route and to use alternate airports
3	Distance to hazardous weather	Distance from the aircraft to areas of ≥30 dBZ precipitation Subjective ratings of how easy it was for pilots to determine the location of hazardous precipitation
4	Weather situation awareness	Pilot perception of weather along the route of flight captured by the pilot's relay of information to the second pilot (communication counts) The relay of weather information gathered from the portable weather application (for the experimental group).
5	Weather display interaction	Interactions with the portable weather application
6	Cognitive engagement	Blood oxygenation changes captured by the fNIR system Subjective workload ratings

Note. fNIR = functional near-infrared.

WSA. A goal of the present study was to assess how the use of portable weather applications affects pilot WSA. We define WSA as a pilot's combined perception of time, current weather distribution along the planned route and alternative routes, areas free of hazardous weather, and weather locations in the near future and the use of alternative routes to avoid hazardous weather. As found by Ahlstrom and Suss (2015), a high WSA implies that a pilot is cognizant of and prepared for weather state changes and will therefore have more time to take appropriate action. During the scenarios, we logged all communications between the pilot and the pilot following and coded each communication message in one of five communication categories.

Weather display interaction. During each simulation run, we recorded all pilot weather display interactions and display durations of individual weather elements.

Cognitive engagement. In this study we were interested in assessing the difference in pilot cognitive engagement between the experimental group (weather display) and the control group (no weather display) as measured by fNIR. Potentially, the weather display can affect a pilot's cognitive load during flight by increasing or decreasing the pilot's cognitive engagement in planning and decision making as indicated by increasing or decreasing oxygenation levels. During the scenario, we measured pilots' prefrontal oxygenation at 2 Hz. In addition to the objective fNIR data, we also captured pilots' subjective workload by having pilots rate their mental workload at the completion of the scenario (7-point scale; 1 = *very low*, 7 = *very high*).

Experimental Design

The human-in-the-loop simulation was conducted as a between-subjects design whereby half of the pilots (experimental group) were equipped with a handheld weather application and the other half (control group) flew without a weather application. Each pilot flew two different scenarios that were counterbalanced across participants.

Procedure

Following an overview of the flight plan and scenario weather conditions (including a preflight weather briefing), the research staff briefed the pilot on the particulars of the flight simulator and how to use the autopilot. If the pilot belonged to the experimental group, he or she was also given a briefing on how to use the weather application, a detailed briefing on the weather information content, and hands-on practice on how to access all the weather information. Thereafter, each pilot flew a 20-min practice scenario with a staff pilot in the copilot seat. During this practice flight, the pilot was asked to perform autopilot maneuvers and to exercise the portable weather application. The staff pilot answered questions and guided the participant as he or she operated the weather application. Pilots were asked to access all weather data elements on the weather presentation and to explore the zoom capability. This practical weather application training continued until the staff pilot was assured that the participant fully understood all the information content and could demonstrate the use of the weather presentation.

Before the start of the test scenario, the pilot was fitted with the fNIR equipment. Next, the pilot was instructed that he or she was going to use VFR and that he or she had planned the route the previous day. The pilot was also instructed that he or she was part of a two-aircraft team ferrying an aircraft to the destination airport. This instruction required the pilot to relay (via radio) weather information and flight decisions to the second pilot (simulated) following their aircraft. Pilots were also instructed that they had to use the autopilot while flying. They were not bound to their preplanned route; they could deviate, turn around, or land at any suitable airport to assure safe alternatives around hazardous weather. Finally, pilots were instructed to obey the Federal Aviation Regulation (FAR) to the best of their ability.

Data Analysis

Traditionally, human factors researchers have used the null hypothesis significance testing (NHST) framework when analyzing their data. However, the NHST framework has many underlying assumptions, and the method is not always straightforward for statistical inference (Dienes, 2011; Gigerenzer, 1998, 2004; O'Keefe, 2003; Wagenmakers, 2007; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). In recent years, many researchers have started using modern Bayesian analysis (Kruschke, 2015; Kruschke, Aguinis, & Joo, 2012) as the Bayesian framework is very flexible and has relatively few methodological constraints. For example, a Bayesian analysis does not rely on *p* values. Bayesian methods are symmetric; they can be used to both reject and accept null outcomes. Bayesian analysis can also equally well accommodate small-*N* samples and large, complex, real-world data sets even though there are missing data or the data sets are unbalanced. When using a Bayesian framework, there is also no need to make corrections for multiple tests on the same data. Using a Bayesian framework, the researcher can also use the outcome of a previous study as prior information for subsequent studies.

In this paper, we are using Bayesian estimation with Markov chain Monte Carlo (MCMC) sampling to determine the posterior distribution of predicted means, standard deviations, and effect sizes as outlined in Kruschke (2015). In the first step of an analysis, we need to determine the scale of our data as recorded by our dependent variables. For example, some data variables are measured on a continuous metric scale (e.g., time in milliseconds), whereas others can be counts (e.g., the number of correct responses) or ordinal ratings (e.g., 1-to-7 rating scale). Next, we need to define the appropriate Bayesian model for each analysis. Because of space constraints, we omit model descriptions here, but readers will find a detailed description of all the Bayesian models used in Kruschke (2015). The model selection is an important step as the model serves as a description of the data. Along with the model selection, we also need to specify a prior distribution on the parameters. For all analyses in the present paper, we use model priors that are vague and noncommittal on the scale of the data, which means that the prior distributions have little effect on the posterior distribution.

Next, we use Bayesian inference by following Bayes' rule: $p(\theta \mid D) = p(D \mid \theta)\, p(\theta) / p(D)$, where the posterior distribution, $p(\theta \mid D)$, is the result of the likelihood, $p(D \mid \theta)$, times the prior, *p*(θ), divided by the evidence, *p*(D). The posterior is our strength of belief in the parameter values and model structure after the data are taken into account. The likelihood is the probability that the data could be generated by the model with parameter values θ. The prior is the strength of our belief in θ before we have taken the data into account. The evidence is the probability of the data according to our model.

In the current framework, the posterior distribution is approximated by a large representative sample of parameter values from the MCMC sampling. Once we have a large sample of representative parameter values, we can assess the mean of a parameter distribution or the differences between values of different parameters. Here, we use a separate decision rule to convert our posterior distributions to a specific conclusion about a parameter value. When plotting the posterior distribution, we include a black horizontal bar that represents the 95% high density interval (HDI). Values inside the HDI have a higher probability density compared to values that fall outside the HDI. When we compare conditions (i.e., perform contrasts), we compute differences at each step in the MCMC chain and present the result in a histogram with the HDI. These histograms show both credible differences and the uncertainty of the outcome. If the value 0 (implying zero difference) is not located within a 95% HDI, we say that the difference is credible. If the 95% HDI includes the value 0, the difference is not credible, as it implies that a difference of zero is a possible outcome. We are also using a region of practical equivalence (ROPE) for effect sizes. The ROPE contains values that, for all practical purposes, are the same as a null effect. If the 95% HDI falls completely within the ROPE margins, we can declare and *accept* the presence of a null effect.

For the analyses in the present paper, we used JAGS (Just Another Gibbs Sampler; Plummer, 2003, 2011), which is a program for analysis of Bayesian hierarchical models using MCMC sampling. We connected to JAGS from a program called R (R Development Core Team, 2011), which is a free software environment for statistical computing and graphics, via the package rjags, an interface from R to the JAGS library. All software for the analysis and figure generation was adapted program code from Kruschke (2015). For the analyses, we used 1,000 steps to tune the samplers and 2,000 steps to burn in the samplers, while running three chains and saving every step in the chain (i.e., we used no thinning). To derive the posterior distributions, we used 200,000 samples.

RESULTS

In the following, we present the results from our dependent measures. However, due to technical problems, some data were not recorded. For the analysis of altitude, data from two pilots in the experimental group were missing (5.8%). For the deviation analysis, data from two pilots in the experimental group (5.8%) and one pilot in the control group (2.8%) were missing. For the analysis of distance to weather, data from three pilots in the experimental group (9%) and one pilot in the control group (2.8%) were missing. Finally, for the fNIR analysis, data from one pilot in the experimental group (2.8%) were missing.

Weather Avoidance

During each scenario flight, we recorded 1,200 altitude measures and 120 deviation measures per pilot. The groups' average altitude and route deviations as a function of the scenario time are illustrated in Figure 5. For the analysis, we averaged each pilot's altitude and deviation distances and used one mean value per pilot using a Bayesian model (Kruschke, 2015) for a metric-predicted variable (i.e., altitude in feet and deviation distance in nautical miles) for two groups (experimental vs. control).

The altitude analysis showed no credible difference between the control group and the experimental group. The mean altitude changes were very similar, with a mean posterior altitude mode of −1,220 ft. for the control group and −1,150 ft. for the experimental group.

As we can see from the deviation data in Figure 6 (also illustrated in Figure 5), the experimental group had larger deviations from the preplanned route than the control group. The posterior means have modes of 1.82 and 0.592 nmi for the experimental and the control group, respectively. The difference of means (mode = 1.18)

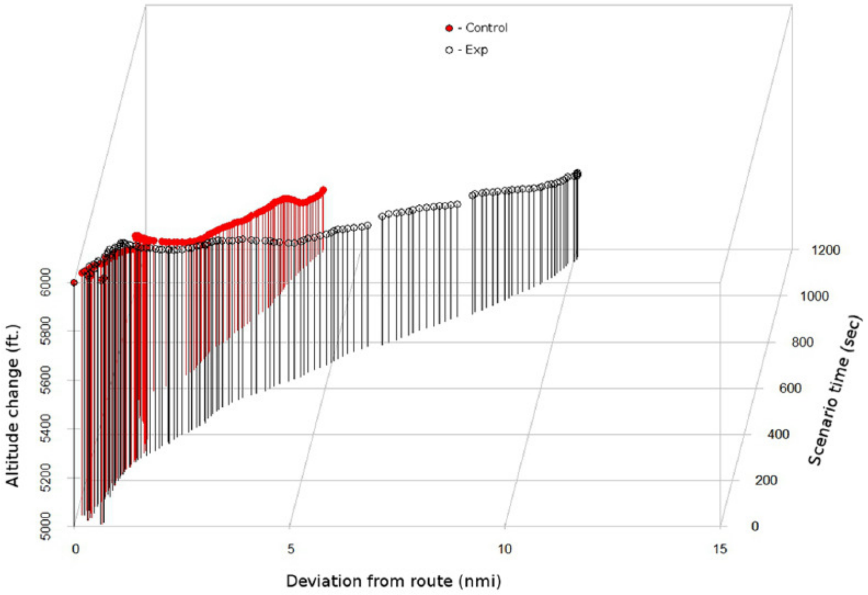


Figure 5. The flight profiles for the experimental group and the control group with mean altitude (feet) and deviations from the preplanned route (nautical miles) as a function of scenario time (seconds).

was credible, as the value 0 was not included in the 95% HDI. The effect size for the deviations was also credible, with a mode of .866 and the value 0 outside the 95% HDI.

Decision Making

Out of 34 pilots in the control group, only one pilot decided to divert to an alternate airport. Of the 36 pilots in the experimental group, six decided to divert. The single pilot in the control group announced the decision to divert at 19 min into the scenario. The mean decision time for diverting for the six pilots in the experimental group was 12.5 min (*SD* = 4.8). Because of the small total number of deviations per group, this difference was not credible.

Distance to Hazardous Weather

For the distance-to-weather analysis, we used only data from 10 min to the end of the scenario as pilots were too far away from the relevant precipitation areas at scenario startup. The average distance-to-weather (along with the average altitude) data for the experimental group and control group are shown in Figure 7. In the figure, we can see the average distance to the weather cell encountered at the end of the scenario. On average, the experimental group kept greater distances to weather than the control group. For the analysis, we first averaged the 10 distance measures for each participant per flight and used one mean value per participant for the analysis, using a Bayesian model (Kruschke, 2015) for a metric-predicted variable (i.e., distance in nautical miles) for two groups (experimental vs. control).

Figure 8 shows the result of the distance-to-weather analysis. On average, the experimental group kept larger distances to hazardous weather cells than the control group. The posterior mean distance for the experimental group had a mode of 9.41 nmi; for the control group, the posterior mode was 6.93 nmi. This difference of means was credible (mode = 2.52) as the value 0 was outside the 95% HDI. Finally, there was a credible effect size with a mode of 0.878.

When asked to rate how easy it was to determine the location of severe precipitation areas at the completion of the scenario, pilots in the experimental group gave higher subjective ratings than the control group. For the analysis, we used a model by Kruschke (2015) for an ordinal

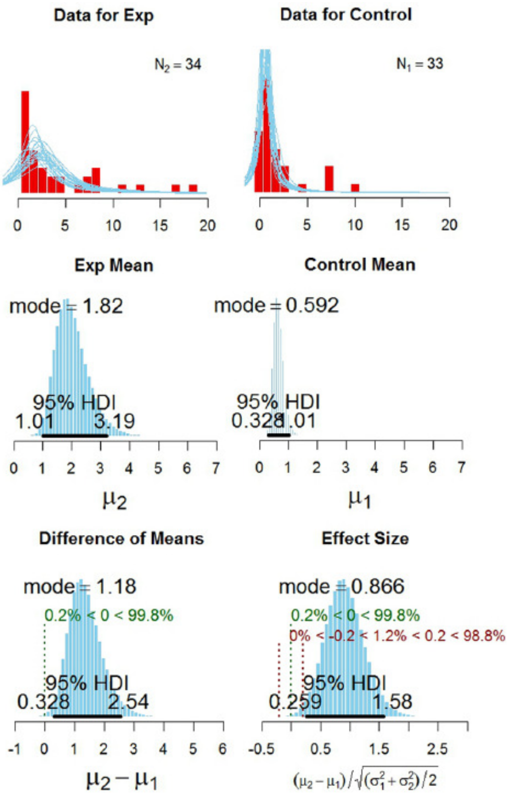


Figure 6. Deviation data (top), posterior distributions for means (middle), difference of means (bottom left), and effect size (bottom right) for the comparison of route deviations between the experimental group and the control group.

predicted variable (i.e., questionnaire ratings) comparing two groups. The outcome showed predicted rating means with modes of 5.09 and 4.34 for the experimental and the control group, respectively. However, the difference of posterior means was not credible, with a mode of 0.702 (95% HDI from −0.27 to 1.65) and the value 0 included in the 95% HDI.

To summarize, when analyzing the vertical flight profiles, we found no credible differences in altitude changes between the experimental group and the control group. However, the analysis of the horizontal flight profiles showed the experimental group to have credibly larger deviations from the preplanned route compared to the control group. The experimental group also kept greater distances to hazardous weather than the control group. However, both groups flew much closer to hazardous weather than what is recommended in current guidelines. Only six participants in the experimental group and two participants in the control group had distances to weather that exceeded 20 statute miles.

WSA

During the scenario flights, pilots were required to relay weather information and flight decisions to the second pilot via radio. We coded each weather and flight communication message with one of the five categories shown in Table 4.

The first category in Table 4, "weather data," captures all communication related to providing weather information, like METARs, TAFs, and the relay of weather state changes acquired from the portable weather application. The second category, "weather direct view," captures relayed weather information acquired from the out-the-window view. The third category, "ground view," captures relayed information associated with terrain, landmarks, and airfields. The category "maneuver/course change" captures communications of decisions to maneuver the aircraft, diverting, and changing course. Last, the category "other" encompasses relayed information about position, heading, altitude, intent, and other nonspecific reports.

Figure 9 (left) shows the predicted weather data communication counts per pilot for the experimental group and the control group. Because the analysis involves a predicted value that is a count (i.e., the number of communications), we used a model by Kruschke (2015) for analysis of data on a count-valued measurement scale. The predicted posterior communication count for the experimental group had a mode of 2.3 (observed mean = 2.3), whereas it was 0.1 for the control group (observed mean = 0.1). As is shown by the posterior distribution to the right, the difference in the weather data communication counts was credible as the value 0 was not included in the 95% HDI.

Figure 10 (left) shows the predicted counts per pilot for the relay of weather direct-view information for the experimental group (observed mean = 9.5) and the control group (observed mean = 11.0). There was a higher predicted count of weather direct-view communications for the

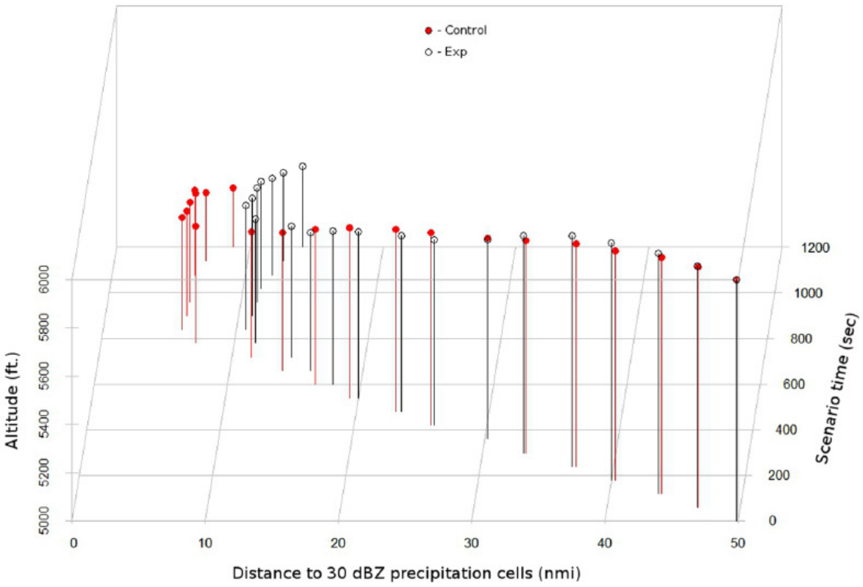


Figure 7. Mean distance to weather (≥30 dBZ precipitation cells) for the experimental group and the control group as a function of mean altitude (feet) and scenario time (seconds).

control group. However, as is shown in the posterior histogram (right), the 95% HDI for the difference of −1.42 includes the value 0. Therefore, we cannot unequivocally state that the difference was credible since a difference of 0 was possible albeit with a very low probability.

Figure 11 (left) shows the predicted counts per pilot for the relay of other information for the experimental group (observed mean = 8.2) and the control group (observed mean = 9.8). There was a higher predicted count of other communications for the control group. As shown by the posterior distribution to the right, the difference in the other communication counts was credible as the value 0 was not included in the 95% HDI.

There were no credible differences between groups in the predicted communication counts per pilot of ground view information (experimental mode = 0.5, control mode = 0.5) and maneuver/course change information (experimental mode = 4.9, control mode = 4.8).

To summarize, the experimental group (using a portable weather application) provided credibly more communications of weather data information than the control group. This finding supports our hypothesis that using a portable weather application will result in an increased WSA. We know from the communications that pilots in the experimental group were aware of weather conditions along the planned route and alternative routes. They were also aware of areas that were free of hazardous weather. The control group provided a higher count of relays of weather direct-view information and a credibly higher count of other information than the experimental group. This outcome is not surprising since the control group relied entirely upon "see and avoid" for hazardous weather avoidance.

Weather Display Interaction

During the simulation, we recorded all pilot interactions with the portable weather application. From these data, we analyzed how often each weather element layer and weather feature was displayed during the scenario. For the result presentation, we analyzed the display time separately for weather layers and weather features. An analysis of the weather layer selections showed that precipitation information accounts for the majority of the total layer display time (i.e., 16 min) with 11.4 min (56%). Besides precipitation, pilots displayed ceiling information and visibility information for 2.3 min (12%) and 1.3 min (6%) of the total time, respectively. The display time was shorter for weather

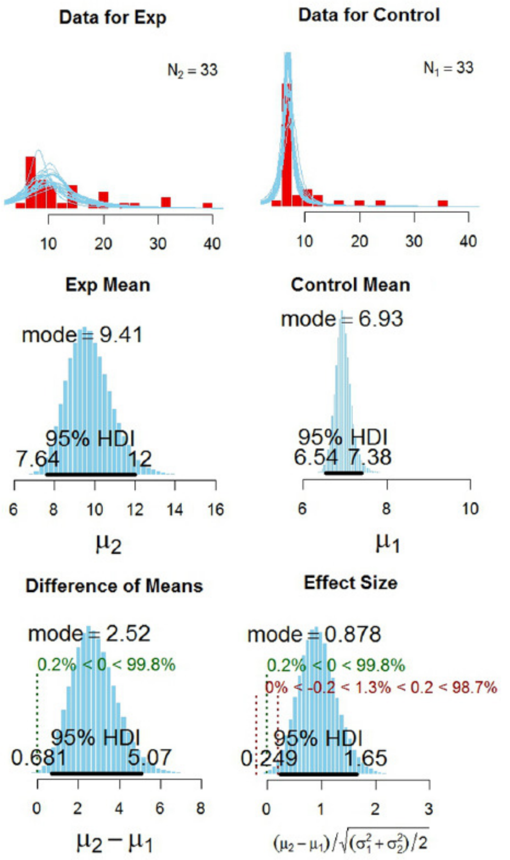


Figure 8. Average distance-to-weather data (top), posterior distributions for means (middle), difference of means (bottom left), and effect size (bottom right) for the comparison of the 10- to 20-min scenario distance to weather (≥30 dBZ cells) between the experimental group and the control group.

information layers, like satellite (1.2 min, 0.06%), turbulence potential (0.07 min, 0.3%), wind (0.3 min, 1.6%), flight category (0.3 min, 1.6%), temperature (0.04 min, 0.2%), and icing probability (0.1 min, 0.8%). The relative-humidity layer was never displayed by any of the pilots. For about 4 min of the total simulation time, pilots viewed only the aircraft position symbol and the preplanned route without displaying any of the weather layers. An analysis of weather information features showed that METARs were displayed for 5.8 min (29%), PIREPs for 3.8 min (19%), and TAFs for 1.5 min (8%) of the total feature display time (i.e., 11.2 min).

Cognitive Engagement

Figure 12 shows the average oxygenation change during flight for the experimental group and the control group. First, the experimental group had, on average, higher oxygenation values than the control group. Second, there was a trend in the oxygenation data where the oxygenation values increased with an increasing scenario time.

For the analysis, we averaged the oxygenation values across all 16 channels and computed an average for each participant. We then used one average oxygenation value per pilot for the analysis using a Bayesian model (Kruschke, 2015) for a metric-predicted variable (i.e., oxygenation values) for two groups. Figure 13 shows the posterior distributions from the analysis. The mean oxygenation for the experimental group (mode = 2.51) was higher than the mean oxygenation for the control group (mode = 1.4). The difference of means was credible, with a mode of 1.14 and the value 0 outside the 95% HDI. The effect size was also credible, with a mode of 0.727 and the value 0 outside the 95% HDI.

This analysis showed a credibly higher oxygenation for the experimental group compared to the control group. Because the experimental group was more active in avoiding weather, we interpret the increased prefrontal blood oxygenation as symptomatic of an increased cognitive engagement due to flight planning and decision making. The control group had no access to weather information while piloting and therefore no ability to plan ahead but instead was limited to the moment-by-moment use of out-the-window information. The increased oxygenation for the experimental group found here is similar to what was found by Ahlstrom and Suss (2015) for pilots who detected METAR symbol changes during flight. In their study, the METAR detections led to an increased level of planning and decision making on part of the pilot compared to pilots who failed to detect the METAR changes.

When asked to rate their mental workload during flight at the completion of the scenario, pilots in the experimental group and the control group gave very similar ratings. An analysis of the ordinal rating data showed posterior means with modes of 4.17 and 4.51 for the experimental and the control groups, respectively. The

TABLE 4: Communication Categories

	Weather Data	Weather Direct View	Ground View	Maneuver/Course Change	Other
Examples	Providing METAR, TAF information, etc. Describing weather on the portable application Reporting weather state changes	Precipitation Visibility Ceiling Unusable altitude Clear of weather Avoiding weather Encountering weather VFR conditions Loss of VFR conditions Weather in sight Cloud location	Report terrain Landmark in sight Landmark not found Airfield in sight Airfield not sighted	Diverting to alternate airfield Increasing/reducing speed Flying direct to (bypassing waypoint) Adding waypoint Turning left/right Turning to heading Climbing Descending Leveling	Position Heading Course Altitude Intent On/off course Navigation problem Nonspecific reports
Experiment group communications	76	313	15	162	270
Control group communications	3	351	16	153	312

Note. METAR = aviation routine weather report; TAF = Terminal Area Forecast; VFR = visual flight rules.

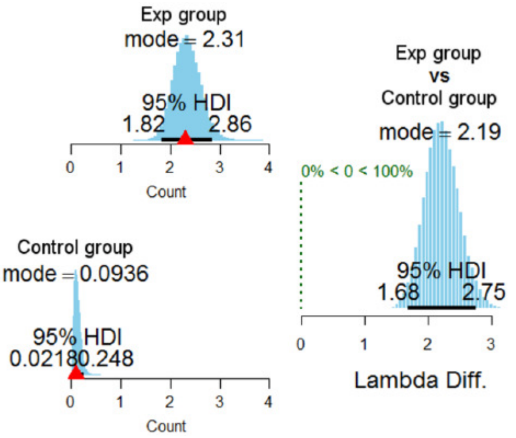


Figure 9. Posterior distributions (left) for the predicted counts of weather data communications per pilot for the experimental group and the control group. The triangle at the bottom of the histogram indicates the mean observed count calculated across the pilots in each group. The histogram to the right shows the posterior contrast for the comparison between the experimental group and the control group.

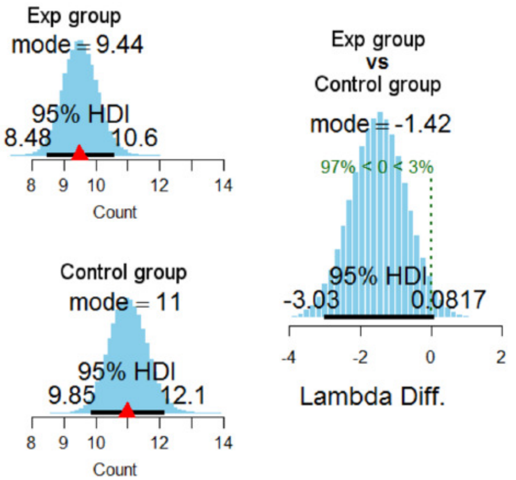


Figure 10. Posterior distributions (left) for the predicted counts per pilot of weather direct-view communications for the experimental group and the control group. The histogram to the right shows the posterior contrast for the comparison between the experimental group and the control group.

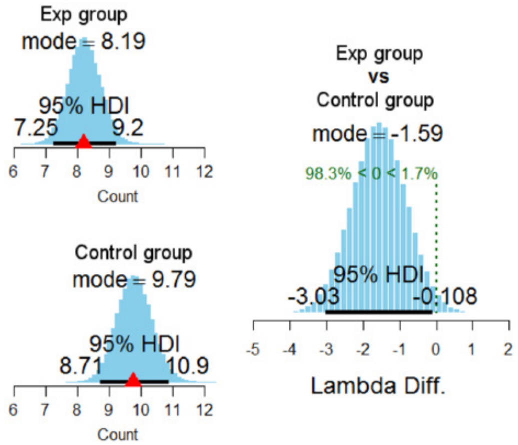


Figure 11. Posterior distributions (left) for the predicted counts per pilot of other communications for the experimental group and the control group. The histogram to the right shows the posterior contrast for the comparison between the experimental group and the control group.

difference of means was not credible, with a mode of −0.33 (95% HDI from −1.18 to 0.462) and the value 0 included in the 95% HDI. This finding is at odds with the objective fNIR recordings, which show a credibly higher oxygenation for the experimental group.

DISCUSSION

In this section, we discuss the study outcome in relation to human factors issues, training, and weather display development. First, we discuss GA pilot WSA from the use of portable weather displays. Second, we discuss the potential benefits of portable weather displays for commercial GA pilots. Third, we discuss the concept of mental workload based on fNIR recordings and subjective postrun workload ratings. Finally, we conclude this section with a discussion of pilot plan-continuation errors and the need to optimize the display of weather information.

The outcome of this study provides empirical evidence that portable weather applications increase pilot WSA. The communication records show that pilots in the experimental group were cognizant of the weather conditions along the preplanned route and at airports. However, pilots also need to make an operational assessment (using their WSA) of what the weather conditions imply for their flight. Thus, there is a link between WSA, an operational assessment, and action. Although the WSA plays a key role in this perception-action

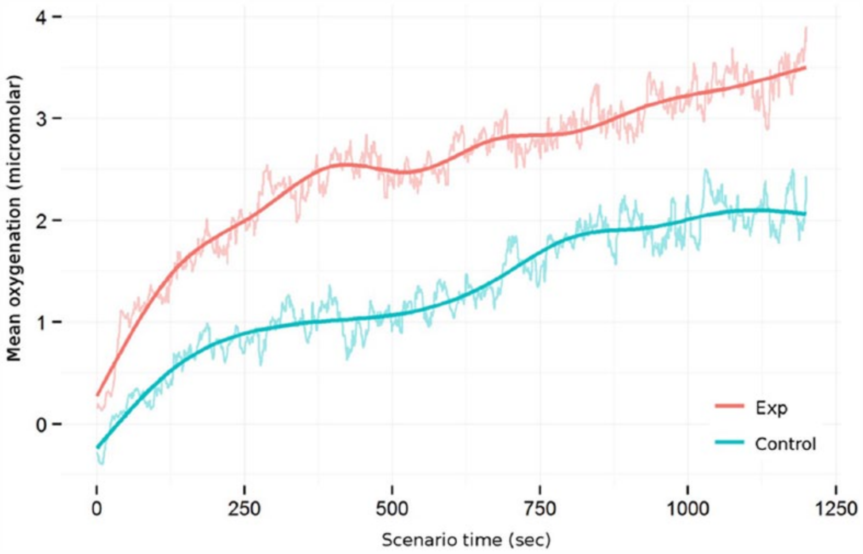


Figure 12. Mean oxygenation data for the experimental group and the control group as a function of scenario time.

chain, it does not guarantee an adequate pilot action. In general, we found that the increased WSA improved flight behavior when avoiding hazardous weather, particularly for decisions to deviate from the route or the ability to stay farther away from hazardous areas. However, we also saw that an increased WSA did not necessarily translate to pilots keeping safe distances to weather; although the experimental group kept credibly greater distances away from ≥30 dBZ cells than the control group, both groups flew much closer to hazardous precipitation cells than what is recommended in current guidelines. What we would like to see is a more appropriate behavioral response—involving greater deviations from hazardous precipitation areas—based on participants' high level of WSA and the information available on the portable weather application. This outcome also suggests that in addition to portable weather applications, pilots may benefit from training on how to interpret weather presentations and how to translate an increased WSA into enhanced flight decisions. Previous work has shown that training on the interpretation of weather cues is successful in influencing pilots to make earlier deviations when they encounter hazardous weather (Wiggins & O'Hare, 2003). In the present study, 57% of the pilots reported having had no additional weather training beyond basic pilot training.

In addition to private GA pilots flying under VMC, we believe there are several reasons why portable weather applications could benefit commercial GA pilots. First, commercial pilots generally have more weather training than private GA pilots. Second, commercial pilots fly more frequently than private pilots and are more likely to encounter hazardous weather. Therefore, commercial pilots need weather information to develop weather avoidance strategies and to calculate the flight risks associated with route selections. Third, commercial pilots frequently operate at night, when weather conditions may be difficult or impossible to detect from the out-the-window view. Fourth, commercial pilots operate under IMC and would therefore benefit from receiving an overview of hazardous weather along the route of flight, at the destination airport, along alternate routes, or at alternate airports. Thus, we believe that commercial airline pilots could benefit from the use of portable weather applications.

Currently, in addition to text-based weather information provided by the Aircraft Communications Addressing and Reporting System, airline pilots have onboard weather radars that provide information about precipitation intensities ahead of the aircraft. To get additional information, airline pilots receive weather updates from ATC and flight dispatch. Potentially, the use of a

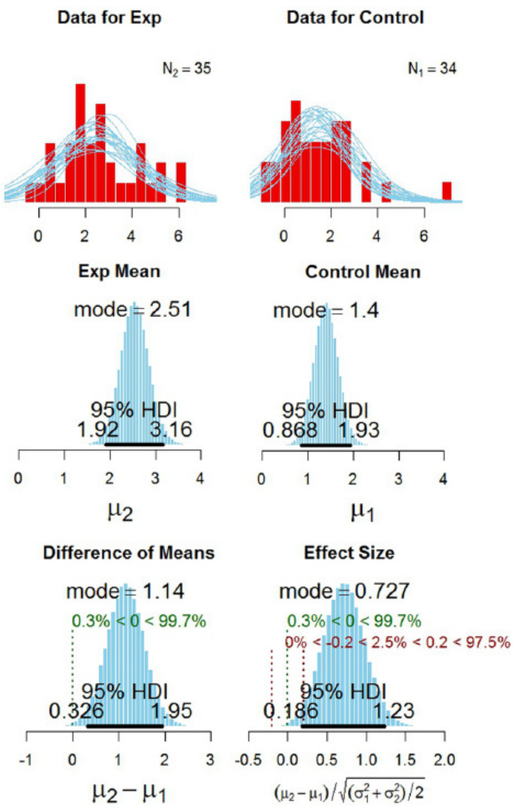


Figure 13. Group oxygenation data (top), posterior distributions for means (middle), difference of means (bottom left), and effect size (bottom right) for the comparison of oxygenation changes between the experimental group and the control group.

portable weather display could provide airline pilots with the opportunity to see a graphical depiction of the big weather picture, which would increase pilot WSA and enhance route negotiations with ATC. Previous research (Gil, Kaber, Kaufmann, & Kim, 2012) uncovered low situation awareness and high cognitive workload for airline pilots deviating around thunderstorms using low automation support. Potentially, the use of a portable weather application could alleviate some of these problems.

In the present study, an objective measure of cognitive engagement revealed heightened activity for the experimental group that lasted the duration of the scenario. Several studies have shown that increased task difficulty corresponds with an increase in activation levels measured via fNIR; for this reason, fNIR activation is often used as an objective estimate of mental workload. In the laboratory, increasingly difficult *n*-back tasks correspond to higher levels of oxygenated hemoglobin and lower levels of deoxygenated hemoglobin in prefrontal cortex as measured via fNIR (Herff et al., 2014). In studies of air traffic controllers, increasing the number of aircraft in a sector led to similar increases in oxygenation (Ayaz et al., 2011; Harrison et al., 2014).

However, authors of previous investigations observed that participants' self-report of workload or effort also increased along with prefrontal blood oxygenation. The present study failed to show a credible difference in subjective reports of workload, although overall the experimental group reported lower workload than the control group, in spite of credibly higher prefrontal blood oxygenation levels. One possible explanation is that subjective postrun workload ratings provide less reliable measures of workload than ratings gathered at fixed time intervals during task performance (Ligda et al., 2010). While fNIR recordings and time interval ratings can show the same trend in the data (Harrison et al., 2014), postrun ratings and fNIR measures might not correlate (as was the case in the present study). Another possible explanation for the discrepancy is that the increased fNIR activation reflects an increase in mental arousal; if so, a common interpretation of the Yerkes-Dodson law dictates that performance would be improved with increasing activation during low arousal or on simple tasks and would hurt during difficult tasks under high arousal (Teigen, 1994; Yerkes & Dodson, 1908).

Improvements in performance due to increasing load have been documented in low-load conditions (Hancock & Warm, 1989; Weiner, Curry, & Faustina, 1984). Additionally, increasing automation during low-workload periods of flight has also been shown to lead to poor pilot performance (Hilburn, Jorna, Byrne, & Parasuraman, 1997). Our finding also shows that an increase in workload or arousal levels is not necessarily associated with poor task performance but may interact with task difficulty. Pilots in the experimental group exhibited higher cognitive engagement than the control group. The increased cognitive engagement resulted in an increased WSA, greater distances from hazardous weather, and an enhanced weather communication. The increased cognitive engagement also improved decision making, such as decisions to deviate from the preplanned route due to weather and decisions to divert to alternate airports.

Despite these positive outcomes, we believe that pilots were not optimizing their use of the portable weather information. During the scenario, pilots mainly used precipitation information and (to a lesser degree) ceiling/visibility information. Pilots also displayed METAR information, PIREPs, and TAFs. Nevertheless, pilots were not able to keep safe distances to weather; they continued their flight toward the destination although it would have been safer to deviate and to land at an alternate airport. This behavior has been described by previous research as plan-continuation errors (Muthard & Wickens, 2003; Wiggins et al., 2014), whereby pilots continue their preplanned flight despite having access to weather information that suggests a deviation or a decision to turn around.

Plan-continuation errors point to a need for optimization of weather display information for single-pilot operations. Pilots in the present study were not optimizing their use of available weather information, and some weather information was barely used at all. This finding is similar to previous research by Burgess and Thomas (2004) and Johnson et al. (2006). As portable display technology and computational power improve, there is a tendency for weather displays to increase in complexity due to an increasing number of weather information elements. Although this complexity might not be a problem for preflight planning, for single-pilot flights, it is likely a problematic trend that will negatively impact the utility of weather displays. Single-pilot operations are highly dynamic. Pilots must integrate what they see out the window with information provided by the instruments and the weather display. We believe this interpretation of weather data, and the necessary realization of what it means for one's flight, is a difficult process that requires a lot of training. However, there is a way to bypass much of the piecewise mental integration and to present information that pilots can act upon directly. Following Runeson's (1977) theory of "smart" perceptual mechanisms that take shortcuts to derive useful information, we believe smart weather display mechanisms could provide useful information to pilots.

For example, instead of solely presenting graphical precipitation intensities that pilots must visually track, interpret, and avoid, smart mechanisms could keep track of the weather information and alert the pilot (Ahlstrom, 2015a; Ahlstrom & Jaggard, 2010). This framework not only presents weather information that pilots can act upon directly, but it also minimizes the likelihood that pilots fail to detect new and updated information. As an added bonus, the use of smart weather display mechanisms might also prevent extensive weather display training that, in the end, will prove unsuccessful because the current weather display framework is not properly tailored to the human perception-action cycle.

CONCLUSIONS AND RECOMMENDATIONS

We found that a portable weather application improves pilots' WSA, cognitive engagement, and weather avoidance maneuvers. However, pilots using the portable application flew much more closely to hazardous precipitation cells (i.e., mode of 5.72 nmi) than what is recommended in current guidelines (i.e., 20 statute miles). Therefore, the use of a portable weather application did not translate to improved flying behavior.

From this outcome, we believe there are three factors that need to be addressed by future research. One is an assessment of the effectiveness of pilot training on how to interpret weather information on modern electronic displays. Another factor that needs further assessment is the potential effect from pilot training on how to translate weather information into enhanced flight decisions. Finally, we recommend research for future weather applications that explores other ways to provide clear display indications of areas to avoid during flight. For example, instead of indicating areas of varying NEXRAD intensities that pilots must interpret, precipitation displays could indicate all areas within 20 statute miles that should be avoided.

ACKNOWLEDGMENTS

The study was sponsored by the Federal Aviation Administration (FAA) Weather Technology in the Cockpit Program Office (ANG-C61). The study plan

was reviewed and approved by the FAA Institutional Review Board.

KEY POINTS

- The use of portable weather applications increases pilot weather situation awareness (WSA) and the ability to avoid areas of hazardous weather.
- Portable weather applications can be used without degrading pilot performance on safety-related flight tasks, actions, and decisions as measured within the constraints of the present study.
- An increased WSA does not automatically translate to enhanced flight behavior.
- There is a need for pilot training on how to interpret weather information on modern electronic displays.

REFERENCES

Ahlstrom, U. (2015a). Experimental evaluation of the AIRWOLF weather advisory tool for en route air traffic controllers. *Aviation Psychology and Applied Human Factors*, *5*, 18–35.

Ahlstrom, U. (2015b). Weather display symbology affects pilot behavior and decision-making. *International Journal of Industrial Ergonomics*, *50*, 73–96.

Ahlstrom, U., Caddigan, E., Schulz , K., Ohneiser, O., Bastholm, R., & Dworsky, M. (2015). *The effect of weather state-change notifications on general aviation pilots' behavior, cognitive engagement, and weather situation awareness* (DOT/FAA/TC-15/64). Atlantic City International Airport, NJ: FAA William Hughes Technical Center.

Ahlstrom, U., & Jaggard, E. (2010). Automatic Identification of Risky Weather Objects in Line of Flight (AIRWOLF). *Transportation Research Part C: Emerging Technologies*, *18*, 187–192.

Ahlstrom, U., & Suss, J. (2015). Change blindness in pilot perception of METAR symbology. *International Journal of Industrial Ergonomics*, *46*, 44–58.

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2011). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage*, *59*, 36–47.

Beringer, D., & Ball, J. (2004). *The effects of NEXRAD graphical data resolution and direct weather viewing on pilots' judgments of weather severity and their willingness to continue flight* (DOT/FAA/AM-04/5). Oklahoma City, OK: Civil Aerospace Medical Institute, Federal Aviation Administration.

Burgess, M. A., & Thomas, R. P. (2004). *The effect of NEXRAD image looping and national convective weather forecast product on pilot decision making in the use of a cockpit weather information display* (Tech. Rep. NASA/CR-2004-212655). Hampton, VA: Langley Research Center.

Coyne, J. T., Baldwin, C. L., & Latorella, K. A. (2005). Influence of graphical METARs on pilot's weather judgment. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 131–135). Santa Monica, CA: Human Factors and Ergonomics Society.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Psychological Science*, *6*, 274–290.

Dutcher, J. W., & Doiron, G. M. (2008, September). *Weather risk management through a systematic approach to the investigation of weather events*. Paper presented at the 39th International Society of Air Safety Investigators (ISASI) Seminar, Halifax, Canada.

Federal Aviation Administration. (2014). *Aviation weather demonstration and evaluation (AWDE) services: Assessment report/results for Mobile MET v2.3*. Unpublished manuscript.

Federal Aviation Administration & National Oceanic and Atmospheric Administration. (1983). *Aviation weather services* (DOT/FAA/AC00-24B). Oklahoma City, OK: Author.

Gigerenzer, G. (1998). We need statistical thinking, not statistical rituals. *Behavioral and Brain Sciences*, *21*, 199–200.

Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*, *33*, 587–606.

Gil, G-H., Kaber, D., Kaufmann, K., & Kim, S-H. (2012). Effect of modes of cockpit automation and pilot performance and workload in a next generation flight concept of operation. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *22*, 395–406.

Goh, J., & Wiegmann, D. A. (2001). Visual flight rules flight into instrument meteorological conditions: An empirical investigation of the possible causes. *International Journal of Aviation Psychology*, *11*, 359–379.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*, 519–537.

Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., . . . Onaral, B. (2014). Cognitive workload and learning assessment during the implementation of a next generation air traffic control technology using functional near infrared spectroscopy. *IEEE Transactions on Human–Machine Systems*, *44*, 429–440.

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., & Schultz, T. (2014). Mental workload during *n*-back task: Quantified in the prefrontal cortex using fNIRS. *Frontiers in Human Neuroscience*, 7(935), 1–9.

Hilburn, B., Jorna, P. G., Byrne, E. A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua & J. Koonce (Eds.), *Human-automation interaction: Research and practice* (pp. 84-91). Mahwah, NJ, Erlbaum Associates.

Hipp, M., Schaub, F., Kargl, F., & Weber, M. (2010). Interaction weaknesses of personal navigation devices. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2010)*, (pp. 129–136). New York, NY: Association for Computing Machinery. doi:10.1145/1969773.1969796

Johnson, N., Wiegmann, D., & Wickens, C. (2006). Effects of advanced cockpit displays on general aviation pilots' decision to continue visual flight rules flight into instrument meteorological conditions. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 30–34). Santa Monica, CA: Human Factors and Ergonomics Society.

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Burlington, MA: Academic Press/Elsevier.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*, 722–752.

Ligda, S. V., Dao, A-Q. V., Vu, K-P., Strybel, T. Z., Battiste, V., & Johnson, W. W. (2010). Impact of conflict avoidance responsibility allocation on pilot workload in a distributed air traffic management system. In *Proceedings of the Human Factors* *and Ergonomics Society 54th Annual Meeting* (pp. 55–59). Santa Monica, CA: Human Factors and Ergonomics Society.

Muthard, E. K., & Wickens, C. D. (2003). Factors that mediate flight plan monitoring and errors in plan revision: An examination of planning under automated conditions. In *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 857–862). Dayton, OH: Wright State University.

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

O'Hare, D., & Stenhouse, N. (2009). Under the weather: An evaluation of different modes of presenting meteorological information for pilots. *Applied Ergonomics*, *40*, 688–693.

O'Hare, D., & Waite, A. (2012). Effects of pilot experience on recall of information from graphical weather displays. *International Journal of Aviation Psychology*, *22*, 1–17.

O'Keefe, D. J. (2003). Colloquy: Should familywise alpha be adjusted? Against familywise alpha adjustment. *Human Communication Research*, *29*, 431–447.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Retrieved from http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Drafts/Plummer.pdf

Plummer, M. (2011). *rjags: Bayesian graphical models using MCMC. R package version 3-5* [Computer software]. Retrieved from http://CRAN.R-project.org/package=rjags

R Development Core Team. (2011). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition*, *7*, 345–376.

Runeson, S. (1977). On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology*, *18*, 172–179.

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, *4*, 525–547.

Uluca, D., Streefkerk, J. W., Sciacchitano, B., & McCrickard, D. S. (2008). Designing automated handheld navigation support. In *Proceedings of the 10th International Conference on Human–Computer Interaction With Mobile Devices and Services* (pp. 481–482). New York, NY: Association for Computing Machinery. doi:10.1145/1409240.1409319

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*, 779–804.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology*, *100*, 426–432.

Ware, C., & Arsenault, R. (2012). Target finding with a spatially aware handheld chart display. *Human Factors*, *54*, 1040–1052.

Weiner, E. L., Curry, R. E., & Faustina, M. L. (1984). Vigilance and task load: In search of the inverted *U*. *Human Factors*, *26*, 215–222.

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human Factors*, *44*, 189–197.

Wiggins, M. W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. *Safety Science*, *65*, 118–124.

Wiggins, M. W., Hunter, D. R., O'Hare, D., & Martinussen, M. (2012). Characteristics of pilots who report deliberate versus inadvertent visual flight into instrument meteorological conditions. *Safety Science*, *50*, 472–477.

Wiggins, M. W., & O'Hare, D. (2003). Expert and novice pilot perception of static in-flight images of weather. *The International Journal of Aviation Psychology*, *13*, 173–187.

Wilson, D. R., & Sloan, T. A. (2003). VFR flight into IMC: Reducing the hazard. *Journal of Aviation/Aerospace Education & Research*, *13*(1). Retrieved from http://commons.erau.edu/jaaer/vol13/iss1/9

Wu, S-C., Gooding, C. L., Shelley, A. E., Duong, C. G., & Johnson, W. W. (2012). Pilot convective weather decision making in en route airspace. In *Proceedings of the ICAS 2012—28th Congress of the International Council of the Aeronautical Sciences*. Retrieved from http://www.icas.org/ICAS_ARCHIVE/ICAS2012/PAPERS/904.PDF

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482. doi:10.1002/cne.920180503

Zimmerman, J. (2013, September). Thunderstorm flying in the age of datalink weather. *Air Facts Journal*. Retrieved from http://airfactsjournal.com/2013/09/thunderstorm-flying-in-the-age-of-datalink-weather/

Ulf Ahlstrom received a PhD in psychology from Uppsala University, Uppsala, Sweden, in 1994. He is currently a research psychologist with the Federal Aviation Administration (FAA) William J. Hughes Technical Center, Human Factors Branch. He has been with the FAA since 1997 following a 2-year postdoctoral fellowship at Vanderbilt University. His current research interests include areas of cockpit weather information and operator workload.

Oliver Ohneiser received his master's degree in informatics from the Technical University of Braunschweig, Germany, in 2011. He joined DLR in 2006 and has worked as a scientific assistant in the Department of Controller Assistance of the Institute of Flight Guidance in Braunschweig since 2010. He does his doctorate in the field of migration tolerant human machine interfaces and develops software applications in the domain of controller assistance systems.

Eamon Caddigan is a human factors specialist at TG O'Brien & Associates, Inc. He received a PhD in psychology from the University of Illinois at Urbana-Champaign in 2012.

Date received: August 28, 2015
Date accepted: March 7, 2016