DOT/FAA/TC-15/42

Initial Assessment of Portable Weather Presentations for General Aviation Pilots

Ulf Ahlstrom, FAA Human Factors Branch
Eamon Caddigan, T. G. O'Brien & Associates, Inc.
Kenneth Schulz, TASC: An Engility Company
Oliver Ohneiser, Institute of Flight Guidance, German Aerospace Center
Robert Bastholm, Spectrum Software Technology, Inc.
Matthew Dworsky, TASC: An Engility Company

September 2015

Technical Report

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the objective of this report. This document does not constitute Federal Aviation Administration (FAA) certification policy. Consult your local FAA aircraft certification office as to its use.

This report is available at the FAA William J. Hughes Technical Center's full-text Technical Reports Web site: http://actlibrary.tc.faa.gov in Adobe® Acrobat® portable document format (PDF).

1. Report No. DOT/FAA/TC-15/42	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Initial Assessment of Portable Weather Presentations for General Aviation Pilots		5. Report Date September 2015	
		6. Performing Organization Code ANG-E25	
7. Author(s) Ulf Ahlstrom, FAA Human Factors Branch Eamon Caddigan, T. G. O'Brien & Associates, Inc. Kenneth Schulz, TASC: An Engility Company Oliver Ohneiser, Institute of Flight Guidance, German Aerospace Center Robert Bastholm, Spectrum Software Technology, Inc. Matthew Dworsky, TASC: An Engility Company		8. Performing Organization Report No. DOT/FAA/TC-15/42	
9. Performing Organization Name and Address Federal Aviation Administration Human Factors Branch William J. Hughes Technical Center Atlantic City International Airport, NJ 08405		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No.	
12. Sponsoring Agency Name and Address Federal Aviation Administration Weather Technology in the Cockpit (WTIC) Program Office 800 Independence Avenue, S.W. Washington, DC 20591		13. Type of Report and Period Covered Technical Report	
		14. Sponsoring Agency Code ANG-C61	

15. Supplementary Notes

16. Abstract

Objective: (a) To examine the potential benefits and effect on pilot flying behavior from the use of portable weather presentations and (b) to assess pilot sensitivity to weather symbology changes. **Method**: Seventy-three General Aviation (GA) pilots volunteered to participate in the study. During simulated flights, participants were randomly assigned either to an experimental group or to a control group and flew a simulated single-engine GA aircraft under Visual Flight Rules (VFR) while avoiding hazardous weather. The experimental group was equipped with a portable "weather application" during flight. We recorded flight profile parameters, Weather Situation Awareness (WSA), decision-making, cognitive engagement, weather-application interaction, and aircraft distance-to-weather. Using a *change-detection* experiment, we assessed participants' sensitivity to symbology changes in portable weather presentations. **Results**: We found positive effects from the use of the portable weather application with an increased WSA for the experimental group. This resulted in credibly larger route deviations and credibly greater distances to hazardous weather (≥ 30 dBZ cells) in the experimental group than in the control group. Nevertheless, both groups flew too closely to hazardous weather compared to what is recommended in current Federal Aviation Administration guidelines. We also found a credibly higher cognitive engagement (prefrontal oxygenation levels) for the experimental group, possibly reflecting increased flight planning and decision-making among the participants. Using a change-detection experiment, we assessed participant discriminability of signal and noise trials using cloud ceiling, precipitation, and PIREP information. We found that discrimination performance was low for all conditions in comparison to the performance of a group of *ideal observers* as measured by the signal detection (SD) metric for discriminability (d'). **Conclusion**: The study outcome supports our hypothesis that the portable weather application can be used without degrading pilot performance on safety-related flight tasks, actions, and decisions. However, it also shows that an increased WSA does not automatically transfer over to improved flight behavior. The outcome shows that participants could learn and operate the portable weather application with relative ease, but training is necessary to help pilots translate weather information into improved flight-behavior strategies. The outcome from the change-detection experiment shows that work is still needed to optimize the symbology for portable cockpit weather presentations. **Applications**: This simulation is part of an initial assessment of the effects of portable weather applications on pilot behavior and decision-making.

17. Key Words Change Detection Cockpit Simulation Cognitive Engagement Portable Weather Application Weather Situation Awareness Weather Symbology		18. Distribution Statement This document is available to the public through the National Technical Information Service, Alexandria, Virginia, 22312. A copy is retained for reference at the William J. Hughes Technical Center Library.	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 85	22. Price
Form DOT F 1700.7 (8-72)	**Reproduction of completed page authorized**		

THIS PAGE IS BLANK INTENTIONALLY.

Table of Contents

List of Illustrations

Tables   **Page**

Acknowledgments

THIS PAGE IS BLANK INTENTIONALLY.

Executive Summary

During this Weather Technology in the Cockpit (WTIC) project, researchers used the Federal Aviation Administration (FAA) William J. Hughes Technical Center (WJHTC) Cockpit Simulator Facility to perform a simulated flight and a *change-detection* experiment on General Aviation (GA) pilot decision-making and Weather Situation Awareness (WSA). Seventy-three GA pilots volunteered to participate in the study. The simulated flight assessed the benefits and effect of the use of portable weather applications on pilot behavior. The change-detection experiment assessed pilot sensitivity to weather symbol changes in weather presentations.

In addition to these main objectives, we also assessed the study outcome in relation to four hypotheses, which state that using a portable weather presentation will improve pilot WSA and assist pilots in avoiding areas of hazardous weather. In the table that follows, we present the four hypotheses and indicate whether they are supported by the study outcome.

Table. Summary of Simulation Hypotheses and Study Outcomes

Hypothesis	Supported by study results	Not supported by study results
H_1: Increased WSA from the use of portable weather applications.	√	
H_2: Using the portable weather presentation results in earlier recognition of weather and weather-state changes, which will afford pilots more time to take appropriate action to avoid adverse weather.	√	
H_3: Earlier adverse weather avoidance decision-making will result in pilots maintaining their appropriate distance from the weather event.		✗
H_4: The portable device can be used without degrading pilot performance on safety-related flight tasks, actions, and decisions.	√	

Hypothesis 1 states that using a portable weather application with selected weather information will result in much higher pilot WSA compared to piloting without the device (i.e., "see and avoid"). We assessed this hypothesis by analyzing flight trajectory and communications data to determine if there was an effect from weather application use on a participant's WSA. When looking at vertical flight profiles, we found no credible differences in altitude changes between the experimental group (portable weather application) and the control group (no weather information). However, when analyzing the horizontal flight profiles, we found a credible difference in route deviation during a convective weather scenario (Scenario A). The experimental group had credibly larger deviations from the pre-planned route compared to the control group. The experimental group had access to information that was used to plan and to make decisions whether to stay on the route or to deviate from areas of hazardous weather. These results indicate a positive effect on participants' WSA when using the portable weather presentation. An analysis of the captured transmissions related to providing weather information from aviation routine weather reports (METARs) and Terminal Area Forecasts (TAFs), as well as information related to weather-state changes acquired from the portable application,

showed that the experimental group provided credibly more communications of weather information than the control group. Assessing the number of deviations to alternate airports and the scenario time at which they had occurred, we found that, compared to the control group, the experimental group made more decisions to divert and that their decisions to divert came earlier in the scenario. However, because of the low numbers, neither the number of decisions to divert nor the time at which the diversions occurred was credibly different between groups. Nevertheless, these outcomes support our hypothesis that participants had an increased WSA when using the portable weather application.

Hypothesis 2 states that using the portable weather presentation results in earlier recognition of weather and weather *state-changes*, which will afford pilots more time to take appropriate action to avoid adverse weather. An analysis of how closely participants flew to areas of hazardous weather showed that the experimental group (using a portable weather application) kept larger distances from hazardous weather cells than the control group. The control group also flew credibly closer to 30 dBZ cells (mode = 2.59 nmi) than the experimental group (mode = 5.72 nmi). However, although the experimental group kept larger distances to hazardous weather than the control group (which supports our second hypothesis), both groups flew closer to hazardous precipitation cells ($\geq$ 30 dBZ cells) than what is recommended in current FAA guidelines (i.e., 20 statute miles).

Hypothesis 3 states that earlier weather avoidance decision-making will result in pilots maintaining a safe distance from the weather event (i.e., following current guidelines). The distance-to-weather analysis showed that the experimental group—using a portable weather application—kept larger distances from hazardous weather cells than the control group. However, the experimental group did not maintain an appropriate distance-to-weather as they flew too closely to $\geq$ 30 dBZ precipitation cells. Therefore, we failed to find empirical support for our third hypothesis.

Hypothesis 4 states that the portable device can be used without degrading pilot performance on safety-related flight tasks, actions, and decisions. To assess participants' cognitive engagement during scenario flights, we recorded prefrontal cortical activity using a Functional Near-Infrared Spectroscopy (fNIR) system. Typically, the fNIR signal from neural activation is a decrease of deoxygenated hemoglobin accompanied by an increase of oxygenated hemoglobin. For Scenario A, we found a credibly higher oxygenation for the experimental group than for the control group. For Scenario B, we found no credible differences between groups. We interpret the increased prefrontal blood oxygenation for the experimental group as symptomatic of an increased cognitive engagement due to flight planning and decision-making. This outcome is similar to what was found by Ahlstrom and Suss (2014) for pilots who detected METAR symbol changes during flight, which led to an increased level of planning and decision-making by pilots. The outcome also supports Hypothesis 4 that the portable weather application can be used without degrading pilot performance on safety-related flight tasks, actions, and decisions.

Finally, we assessed participant sensitivity to weather symbol changes in images from the portable weather application. Using a change-detection experiment, we assessed participant discriminability of signal and noise trials using Cloud Ceiling, Precipitation, and Pilot Report (PIREP) information. In general, participant discrimination performance was low for all conditions in comparison to the performance of a modelled group of *ideal observers*.

For the discrimination of color changes in Cloud Ceiling areas (i.e., color-shaded areas), we found modest discrimination accuracy compared to the performance of the modelled ideal observers, but there were no credible differences in discriminability between image conditions with a high (100%) vs. low (50%) number of color-shaded areas.

For the discrimination of PIREP symbols, neither the manipulation of the number of PIREP symbols (6 vs. 3), the manipulation of the number of precipitation cells (100% and 50%) along with the PIREP symbols nor the manipulation of the PIREP symbol intensity changes along with 100% and 50% precipitation had any effect on discrimination performance.

These findings imply that participants had great difficulty discriminating signal from noise trials and that the manipulation of the 100% and 50% levels were not enough to differentiate performance. The only exception is a credible difference in response times for intensity changes of PIREP symbols between the 100% and 50% precipitation conditions, with longer response times for the 100% precipitation condition. The outcome also implies that work is still needed to optimize the symbology for portable cockpit weather presentations. All symbol and background combinations should provide optimal luminance contrast, thereby enhancing symbol discrimination and reducing the time needed to differentiate all elements in the presentation.

THIS PAGE IS BLANK INTENTIONALLY.

1. INTRODUCTION

In previous research by Ahlstrom and Dworsky (2012) and Ahlstrom and Suss (2014), pilots used large-sized weather presentations (viewing angle subtending 9° horizontally and 20° vertically) mounted in the instrument panel that displayed a fixed number of weather elements. However, today's commercial weather products are often viewed on smaller cockpit installed displays, on Portable Electronic Devices (PEDs), or on mobile devices (Federal Aviation Administration [FAA], 2010). Furthermore, commercial products are often characterized by a large number of interactive weather elements and by an elaborate menu structure. Although mobile devices show promise as a tool for navigation (Ware & Arsenault, 2012), there is insufficient data on the operational benefits from the use of portable weather applications. Assessments are needed to evaluate the impact on pilot Weather Situation Awareness (WSA) as the relatively small viewing angle, large number of elements, and device interaction can pose additional constraints in multitasking situations like those encountered during single-pilot operations. Therefore, the purpose of this study is to perform an initial evaluation of the operational benefits and WSA effect of portable weather applications on General Aviation (GA) pilot behavior.

1.1 Purpose

The goal of this study is to perform an initial human factors assessment of portable cockpit weather applications on GA pilot behavior. Currently, there are no published research data on the use of portable weather applications in the cockpit; therefore, the specific objectives of this study are

- to assess the effect on pilot decision-making and WSA from the use of portable weather presentations.

- to evaluate pilot sensitivity to changes of separate elements in portable weather presentations.

1.2 Simulation Flight Hypotheses

To examine the specific objectives of the study, these four hypotheses were proposed to help better analyze the main hypothesis that the presentation will improve pilot WSA and assist pilots in avoiding areas of hazardous weather.

1. Using a portable device with selected weather information will result in much higher pilot WSA compared to piloting without the device (i.e., "see and avoid").

2. Using the portable weather presentation results in earlier recognition of weather and weather-state changes, which will afford pilots more time to take appropriate action to avoid adverse weather.

3. Earlier adverse weather avoidance decision-making will result in pilots maintaining their appropriate distance from the weather event.

4. The portable device can be used without degrading pilot performance on safety-related flight tasks, actions, and decisions.

2. METHOD

The method and simulation equipment for the present study was the same as the simulation and part-task equipment used in Ahlstrom and Dworsky (2012) and Ahlstrom and Suss (2014). For

the simulation, we used a GA cockpit simulator with a 180° out-the-window-view. For the change-detection experiment we used the same Stimulus Experiment System (SES) and change-detection paradigm as outlined in Ahlstrom and Suss (2014).

2.1 Participants

Seventy-three GA pilots volunteered to participate in the study. The participant pool consisted of commercial, military, and private pilots provided by the FAA William J. Hughes Technical Center (WJHTC). Each pilot participated in a simulated flight and a change-detection experiment. Participants were randomly assigned to experiment (portable application present) and control (no portable application) groups. Table 1 shows the characteristics for each group, and Table 2 shows the ratings for each group. Three participants did not fly in the simulator due to technical issues.

Table 1. Descriptive Characteristics of Study Participants

| | | | | Flight hours accrued | | | | | |
| | | Age (years) | | Total | | Instrument | | Instrument (last 6 months) | |
Group	n	Mdn	Range	Mdn	Range	Mdn	Range	Mdn	Range
Exp	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	0–25,000	5	0–200

Note. Mdn = Median. The median is the numerical value separating the upper half of a data sample from the lower half.

Table 2. Study Participant Ratings

| | Ratings | | | | | | | |
Group	Private	Commercial	ATP	Glider	SEL	SEA	MEL	Airship
Exp	11	11	16	2	24	6	19	0
Control	6	12	20	2	21	2	21	0

	Instrument	CFI	CFI II	MEI	Helicopter	A&P	IA
Exp	20	9	9	6	2	5	1
Control	20	11	14	6	3	4	1

Note. Exp = Experimental group; ATP = Airline Transport Pilot; SEL = Single Engine Land; SEA = Single Engine Sea ; MEL = Multi Engine Land; CFI = Certified Flight Instructor; MEI = Multi Engine Instructor; A&P = Airframe and Powerplant; IA = Instrument Airplane.

2.1.1 Informed Consent Statement

Each participant read and signed an Informed Consent Statement (Appendix A) before starting the simulation flight. The informed consent statement describes the study, the foreseeable risks, and the rights and responsibilities of the participants, including a reminder that participation in the study is voluntary. It also stated that the participant could withdraw from participation at any time without penalty. All the information that the participant provided to us, including Personally Identifiable Information (PII), has protection from release except as required by statute. Signing the form indicated consent and that the participant understood his or her rights as a participant in the study.

2.1.2 Biographical Questionnaire

Participants completed a brief Biographical Questionnaire (see Appendix B). This questionnaire consisted primarily of questions related to pilots' flight experience and previous experience with weather displays.

2.1.3 Post-Scenario Questionnaire

All participants completed a 10-item rating-scale questionnaire (see Appendix C), including an additional yes or no item and an open-answer item to report any discomfort with the functional Near Infrared (fNIR) sensor.

2.1.4 Participant Familiarity with Weather Displays and Weather Interpretation Training

A total of 16 participants had previous experience with weather displays (8 in each group). Of these 16, only 8 had experience with portable weather displays (three in the control group and five in the experimental group). There were only a few participants with prior training of how to interpret weather displays (two in the control group and one in the experimental group).

2.2 Research Personnel

Members of the research team (see Appendix D) set up the simulations and experiments, developed and tested scenarios, prepared the SES for operation, conducted briefings, collected data, and wrote this final report.

2.3 Equipment

2.3.1 GA Cockpit Simulator

The simulation was performed in a GA cockpit simulator configured to simulate a Mooney Bravo single-engine aircraft (see Figure 1). The simulator was an integrated system that comprised a simulator-technician workstation, a cockpit system, and a voice communications system. The simulator was enclosed (fuselage) and equipped with a 180° out-the-window view. The cockpit simulator ran on the Windows 7 operating system using Microsoft Flight Simulator 2004 via the Project Magenta workstation control scheme for a single-engine aircraft. The out-the-window view was generated by the Lockheed Martin Prepare 3D software.

Figure 1. The cockpit simulator.

2.3.2 Cockpit Glass Panel

The cockpit simulator used a glass cockpit panel template (see Figure 2).

Figure 2. The cockpit glass display.

2.3.3 Weather Presentation

During the flight simulation, pilots in the experimental group (half of the participants) were equipped with a portable weather application. Each participant had the portable device (an Apple iPad Air 2) on his or her thigh, secured by a leg strap (see Figure 3). The iPad was running an application developed by the FAA and the National Center for Atmospheric Research (NCAR).

Figure 3. The iPad secured by leg strap.

The specifically designed weather presentation provided a platform for research on GA pilot use of weather and airspace information in the cockpit. The weather information came from the Aviation Digital Data Service (ADDS) website created by the National Weather Service (NWS). The weather application displayed information overlaid on a map with a red aircraft position symbol. A menu allowed layer selections of graphical area depictions of flight-rules categories (e.g., visual flight rules [VFR], instrument flight rules [IFR], ceiling information, visibility, precipitation, icing probability, turbulence potential, wind, temperature, relative humidity, and satellite imagery information, as Figure 4 shows). In addition, the user could display weather information from aviation routine weather reports (METARs), terminal area forecasts (TAFs), and pilot report (PIREP) icons. Pilots could either interpret the icon meanings or tap on them to display additional text-based information.
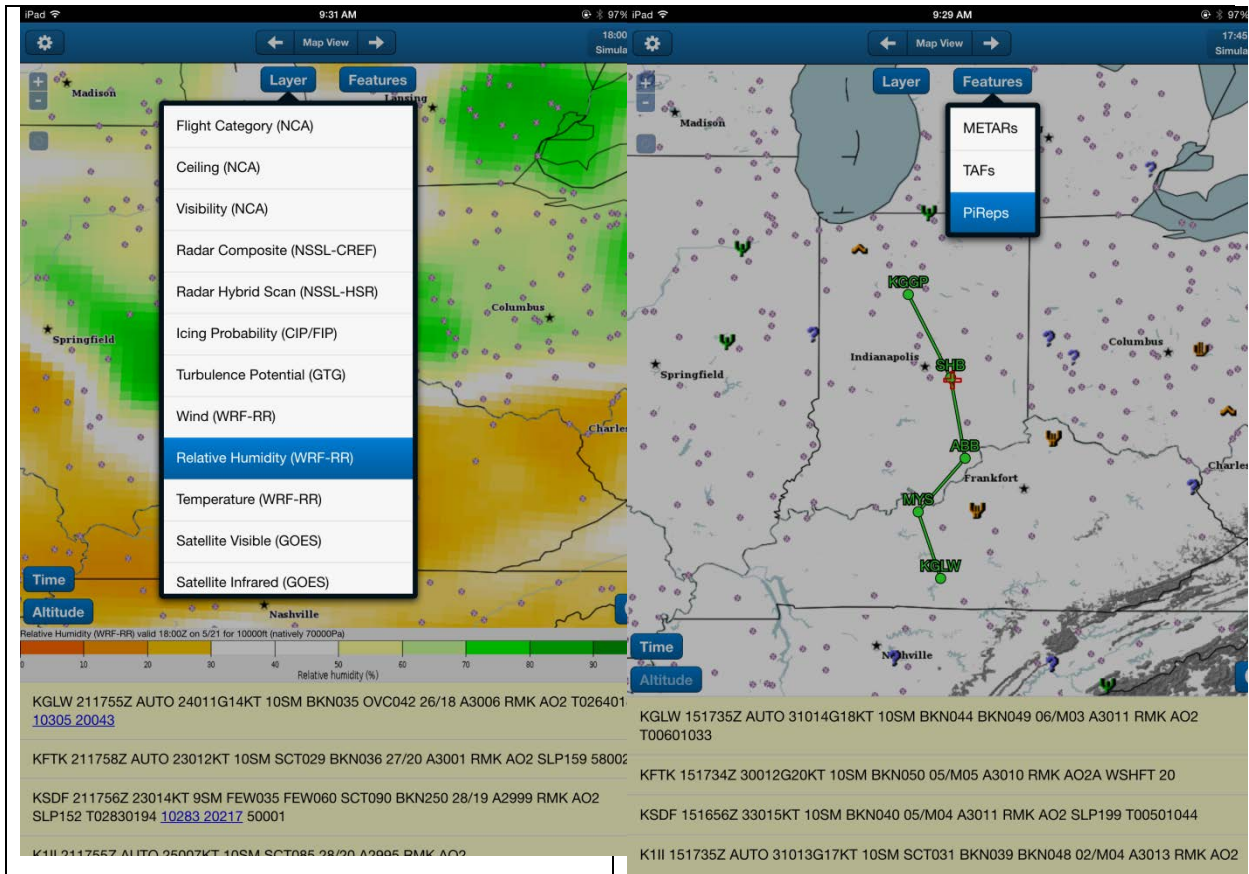
Figure 4. Illustration of the weather application menus.

Users could zoom in or zoom out using zoom gestures or by tapping the "+" and "-" buttons in the left top corner of the display. Figure 5 shows examples of four display images.
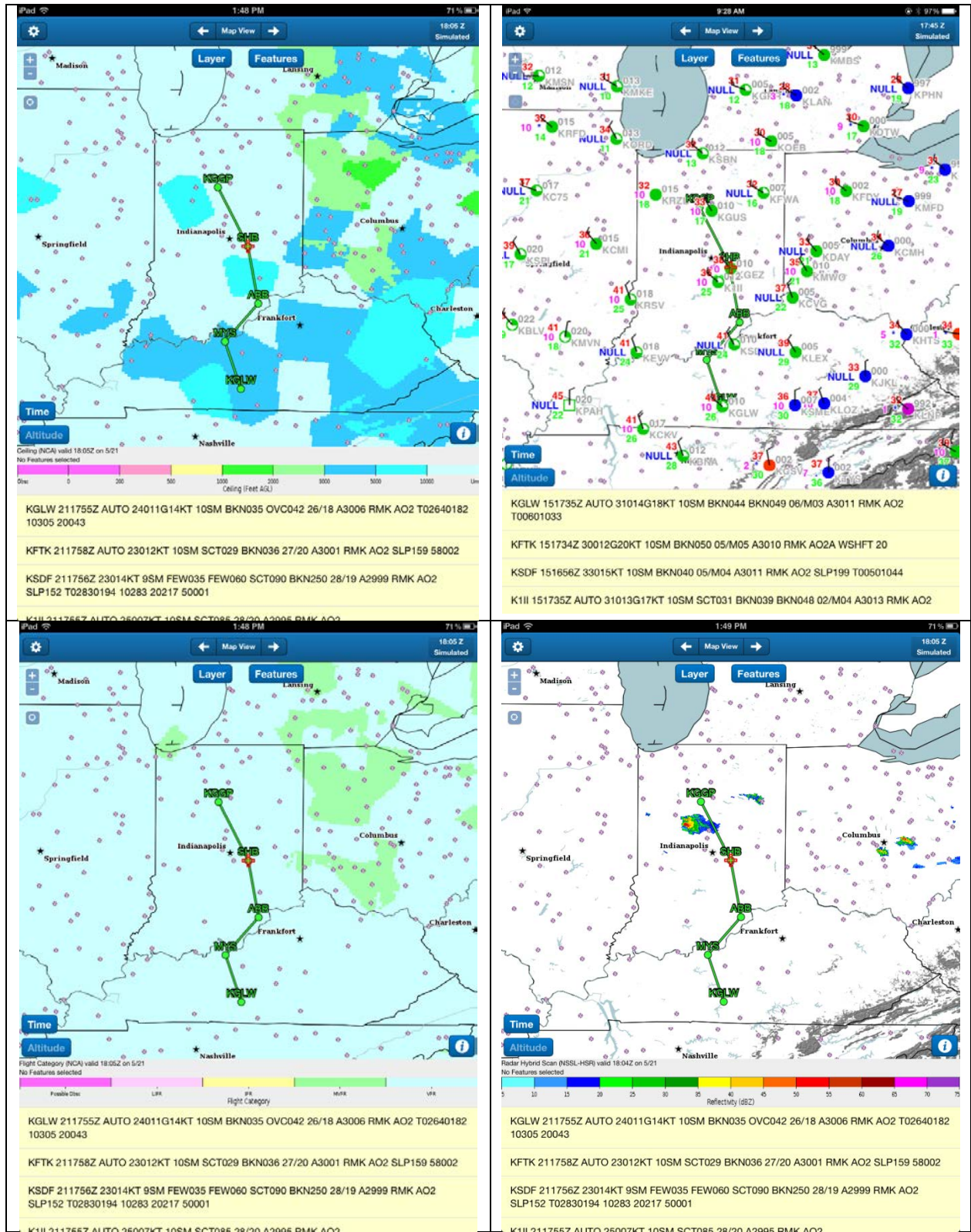
Figure 5. Illustration of the portable weather application with route (in green) and aircraft position symbol (red "plus"); ceiling information (top left), METAR (top right), flight category (bottom left), and precipitation (bottom right) information.

2.3.4 Flight Scenarios

We used three separate VFR scenarios during the simulation. Two of the scenarios were experimental test scenarios (data collection), and a third scenario was a practice scenario (no data collection). Each scenario was 20 minutes in duration. For each scenario, pilots commenced their flight in the cruise phase. During the simulation, we counterbalanced the order of the scenarios across pilots. The experiment group (half of the pilots) had access to a portable weather application during flight while the control group could only acquire weather information from what they saw out-the-window (VFR flight without a portable weather application).

In an en route convective scenario (Scenario A), pilots departed from Glasgow Municipal, Glasgow, KY (KGLW) and encounter thunderstorms along the route of flight to Logansport/Cass County, IN (KGGP). An example of the out-the-window view is illustrated in Figure 6, and the scenario route is illustrated in Figure 7.
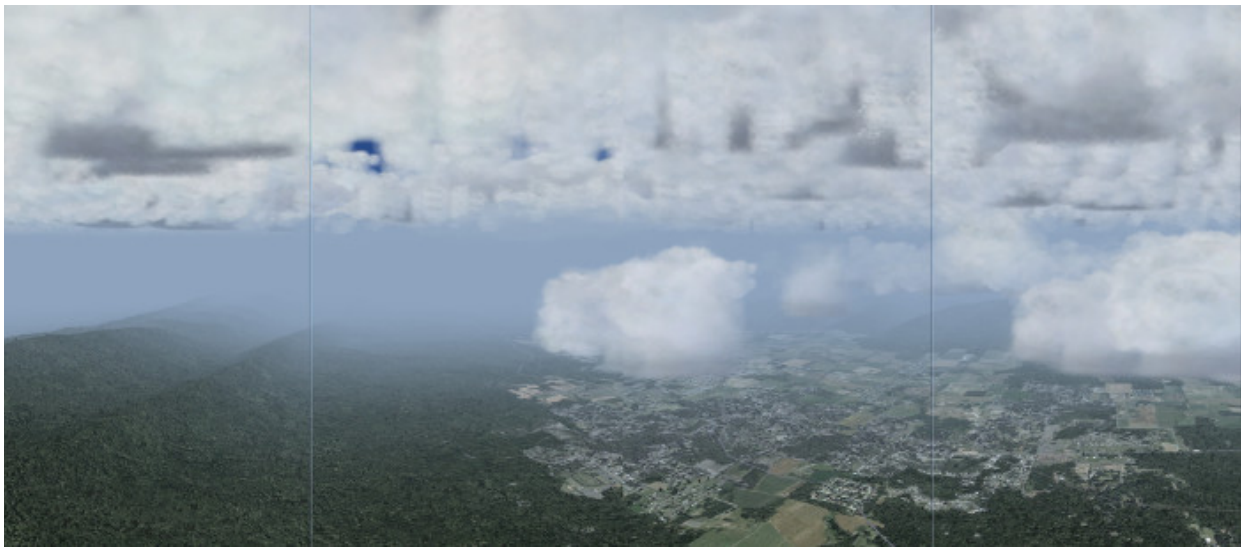
Figure 6. An example of the out-the-window view during Scenario A.
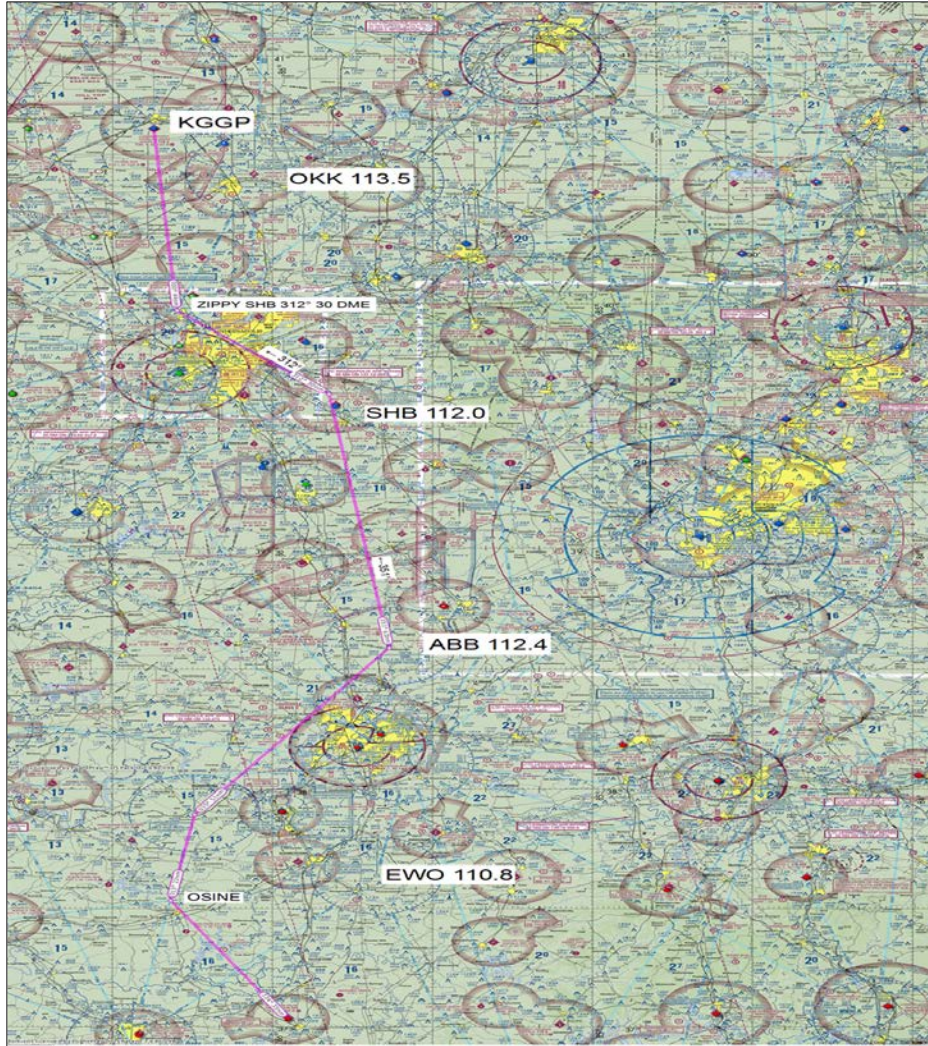
Figure 7. Scenario A (start-up altitude = 6,000 ft) - Depart: KGLW (Glasgow Municipal, Glasgow, KY). Arrive: KGGP (Logansport/Cass County, IN). Route: KGLW (Glasgow, KY) → OSINE, MYS (Mystic, KY) → ABB (Nabb, KY) → SHB (Shelbyville, IN) → ZIPPY → KGGP (Logansport, IN).

In a second ceiling, icing, and visibility scenario (Scenario B), pilots departed from Dayton, OH (KDAY) and flew to Glasgow Municipal, Glasgow, KY (KGLW). At the time of departure, the area surrounding the pre-planned route was VFR. However, low clouds and freezing levels have caused problems with icing along the route with several PIREPs, indicating light to moderate icing. An example of the out-the-window view is illustrated in Figure 8, and the scenario route is illustrated in Figure 9.

Figure 8. The Illustration of the out-the-window view during start-up for Scenario B.
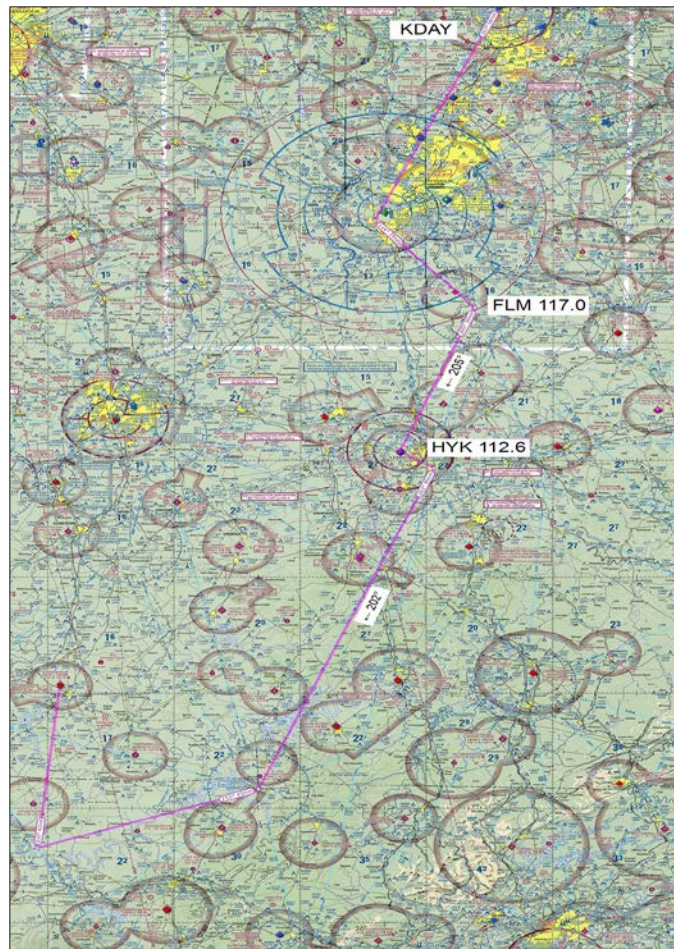
Figure 9. Scenario B (start-up altitude = 2,000 ft) - Depart: KDAY (Dayton, OH). Arrive: KGLW (Glasgow Municipal, Glasgow, KY). Route: KDAY (Dayton, OH) → CVG (Cincinnati, OH) → FLM (Falmouth, OH) → KLEX (Bluegrass, Lexington, KY) → HYK (Lexington, KY) → LVT (Livingston, OH) → HARME → KGLW (Glasgow Municipal, Glasgow, KY).

In our practice scenario (Scenario C) pilots started near Dryer (DJB), having departed from Toledo Executive, Toledo, OH (KTDZ), and flew towards the destination Erie International Airport, Erie, PA (KERI) as illustrated at the bottom of Figure 10. Although there was no precipitation along the route of flight, there was a forecasted area of reduced ceiling and visibility along the route.

Figure 10. Test Scenario - Depart: KTDZ (Toledo Executive, Toledo, OH). Arrive: KERI (Erie Intl, Erie, PA). Route: KTDZ (Toledo Executive, Toledo, OH) → VASHO → SKY (Sandusky, OH) → DJB (Dryer, OH) → JFN (Jefferson, OH) → KERI (Erie Intl, Erie PA).

2.3.5 Stimulus Experiment System

Researchers used the Stimulus Experiment System (SES) software to display the change-detection stimuli. This is the same SES system used by Ahlstrom and Suss (2014) in a previous change-detection experiment. This system allowed the researchers to administer the experimental tasks to the participant by assigning a coded identifier to the participant and then automatically presenting a set number of experimental trials to the participant. Data (i.e., participants' responses) are recorded automatically and then written to a data file. The SES software is installed on Hewlett Packard desktop computers equipped with a Dell P2212H LCD monitor.

Similar to Ahlstrom and Suss (2014), we also used a one-shot change-detection paradigm to assess participants' ability to detect changes between two weather presentation images (i.e., Image #1 and Image #2), as illustrated in Figure 11.
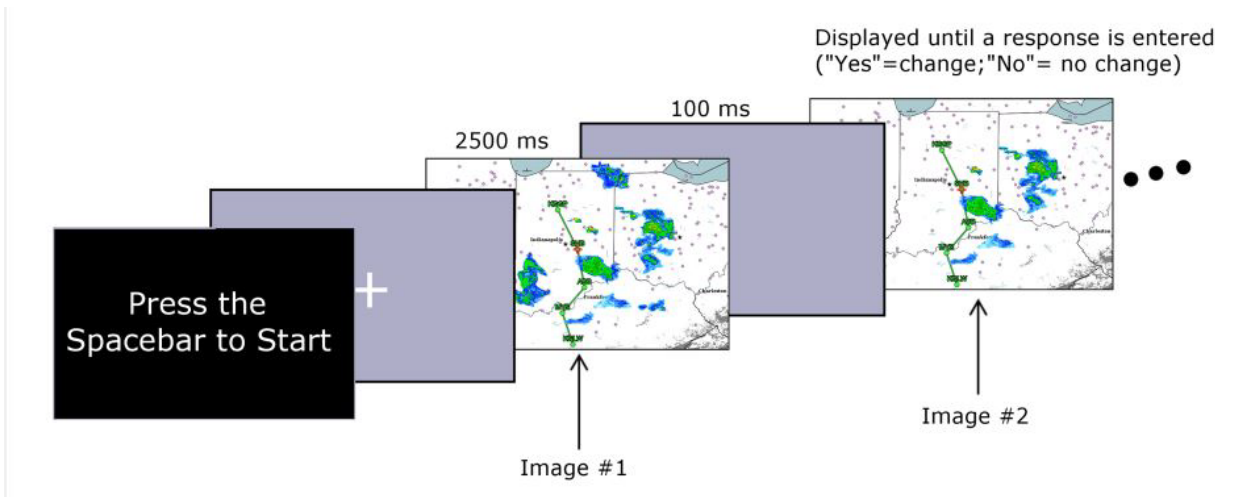
Figure 11. Illustration of the one-shot change-detection technique. Adapted from Rensink, 2002.

2.3.6 Functional Near-Infrared Spectroscopy

To assess pilots' cognitive engagement during simulation runs, the Ahlstrom and Dworsky (2012) study and the Ahlstrom and Suss (2014) study used an objective fNIR recording during simulation flights. In this study, we used the same system during the simulation to record functional cortical activity during flight. The fNIR technology uses specific wavelengths of light to measure changes in the relative ratios of deoxygenated hemoglobin and oxygenated hemoglobin due to brain activity. 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 channels per wavelength. With two wavelengths and dark current recordings for each of the 16 channels, the system generates a total of 48 measurements for each 2 Hz sampling period. The risk associated with using the fNIR sensor is less than the risk associated with spending an equivalent amount of time in sunlight, in the United States, without wearing a hat.

2.3.7 Voice Communication System

The laboratory voice communication system provided a one-way link between the pilot and SME, and research personnel, who played the role of the pilot of the "aircraft following." The pilot's microphone was continuously "live"; no Push-To-Talk (PTT) function was necessary. Pilot communications were recorded digitally as Windows Media Audio (WMA) files. The experimenter made written records of the times and contents of pilot communications for subsequent coding and analysis.

2.4 Procedure

2.4.1 Cockpit Simulator Briefing

Before entering the simulator cockpit, participant pilots received an initial briefing (self-paced PowerPoint presentation) about the flight procedures, scenario, and the simulation system. Participants also received an initial orientation of the portable weather presentation and weather data elements. The SME or the researcher also guided the pilot during a hands-on practice session to allow the pilot to become familiar with using the portable weather presentation. At the end of this briefing, the SME demonstrated the basic aircraft controls and the cockpit console. The SME also instructed each participant about how to change the radio frequency and how to perform autopilot operations. After this briefing, the participant performed a practice scenario flight.

2.4.2 Simulation and Experimental Designs

The flight simulation was conducted as a between-subjects design. Half of the pilots (experimental group) were equipped with a handheld weather application, and the other half (control group) flew without a weather application. The purpose of this design was to allow researchers to assess the effect on pilot decision-making and behavior from the use of a handheld weather application. The simulation and experimental conditions were counterbalanced across pilots, with the constraint that pilots always performed the simulation flight before performing the change-detection experiment.

2.4.2.1 Independent variable

The only independent variable was the availability of a portable weather application.

2.4.2.2 Dependent variables

During the simulation flights, we recorded six dependent variables (as outlined in Table 3).

Table 3. Simulation Dependent Measures

Number	Dependent variable	Description
1	Flight profile measures	The vertical (altitude) and horizontal (deviation from pre-planned route) flight profile.
2	Weather situation awareness (WSA)	Pilot perception of weather along the route of flight captured by the pilot's communication of information to the second pilot.
3	Decision-making	Pilot decision to deviate from the pre-planned route and/or to divert to an alternate airport.
4	Cognitive engagement	The blood oxygenation changes captured by the fNIR system.
5	Weather presentation interaction	Recorded pilot interactions with the portable weather application.
6	Distance to hazardous weather	During the three simulation scenarios, we measured the distance from the aircraft to areas of ≥ 30 dBZ precipitation and reported icing (defined by icing PIREP symbol).

2.4.3 Simulation Data Collecting Procedures

Before the start of each test flight scenario, each pilot was fitted with the fNIR equipment and a communications headset. Pilots were provided with an aeronautical chart, with a plot of the filed route, and with labels identifying and giving the frequencies for the Very High Frequency Omnidirectional Radars (VORs) used as waypoints (as shown in Figure 9). Pilots were instructed to maintain the flight within Visual Meteorological Conditions (VMCs), deviating from the filed route if necessary. They were told that they could elect to divert to an alternate airfield should the situation warrant. Participants were also told that they were part of a two-aircraft team ferrying an aircraft to the destination airport; they were instructed to communicate weather information and navigation instructions to the pilot following.

Once the fNIR system was calibrated and the pilot was ready to fly, the simulation scenario began. At the end of the flight scenario, pilots completed a questionnaire (see Appendix C). After the questionnaire, the research team had a final debriefing with each pilot.

2.4.4 Data Handling Procedures

All information gathered from participants is strictly confidential, and all participants are anonymous. We assigned a coded identifier to each participant. The identifier did not appear on the informed consent statement—because that is identified by the participant's signature. We tagged all other data collection forms, computer files, electronic recordings, storage media, and so on, containing participant information only with the coded identifier—not the name or personal identifying information of the participants. We are retaining original documents, recordings, and files as collected. All data editing, cleanup, and analysis were performed on copies traceable to the original sources.

2.4.5 Data Analysis

We used Bayesian estimation to analyze data from the current study as outlined in Ahlstrom and Suss (2014). This analysis framework uses Bayesian estimation with Markov Chain Monte Carlo (MCMC) sampling to determine the posterior distribution of parameters (e.g., means, scale or standard deviations, and effect sizes). During the analysis, we used JAGS ("Just Another Gibbs Sampler"; Plummer, 2003, 2011) that we called from R (R Development Core Team, 2011) via the package rjags. All software for the analysis and figure generation is adapted program code from Kruschke (2014).

When using the Bayesian analysis framework, we generate a sample posterior distribution, which is a distribution of credible parameter values. We can use this large distribution of representative parameter values to evaluate various parameters—such as means, scales, or effect sizes. We can also compare differences between parameter distributions. 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). The HDI is a very important concept for the forthcoming analyses. This is because every value inside the HDI has a higher probability density compared to values that fall outside the HDI. When we compare conditions, 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 means that a difference of 0 is a possible outcome.

We are also showing a Region of Practical Equivalence (ROPE) in the histograms. 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 for an effect size, we can declare the presence of a null effect, and unlike traditional analyses, we can accept the null outcome. If, on the other hand, the entire ROPE falls outside the 95% HDI, we can reject the presence of a null effect.

For all analyses, we used 1,000 steps to tune the samplers and 2,000 steps to burn-in the samplers, while running 3 chains and saving every step in the chain (i.e., we used no thinning). To derive the posterior distributions, we used 200,000 samples. For all analyses, we use priors that are vague and noncommittal on the scale of the data.

2.4.6 Derivation of Flight Path, Deviation, and Distance-to-Weather Measures

The weather scenarios in the present study contained different weather patterns that might affect pilot behavior. For example, one scenario contained convective activity whereas another scenario contained areas of icing and low ceiling/visibility. We measured the distance from the aircraft location (i.e., latitude/longitude) to the closest point of approach for $\geq$ 30 dBZ precipitation cell intensities (visualized as yellow pixels) at one-minute intervals. We also measured the distance between the aircraft and the area of reported icing along the flight path. For all scenario flights, we also recorded the aircraft's position (in 10-second intervals) relative to the pre-planned route.

Because we used prerecorded weather scenarios, we did not have access to the latitude/longitude position data for each weather element and symbol location. To circumvent this fact, we used C++ to program an automated evaluation tool that loaded and evaluated weather screenshots and recorded log files from the scenarios. 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. This equation serves as a simple alternative to the haversine formula, sufficient for our computations due to the short scenario distances:

$$d = acos(sin(latA) \cdot sin(latB) + cos(latA) \cdot cos(latB) \cdot cos(lonB - lonA)) \cdot 3440.065.$$

During the calculation, the algorithm assessed whether a perpendicular latitude/longitude point was on the defined route segment and whether the great circle distance between the aircraft position and the perpendicular point was smaller than the shortest vector distance. If true, the algorithm updated the location value. This implies that equidistant path deviations were calculated between the blue path and all the hypothetical red aircraft positions (as illustrated in Figure 12).
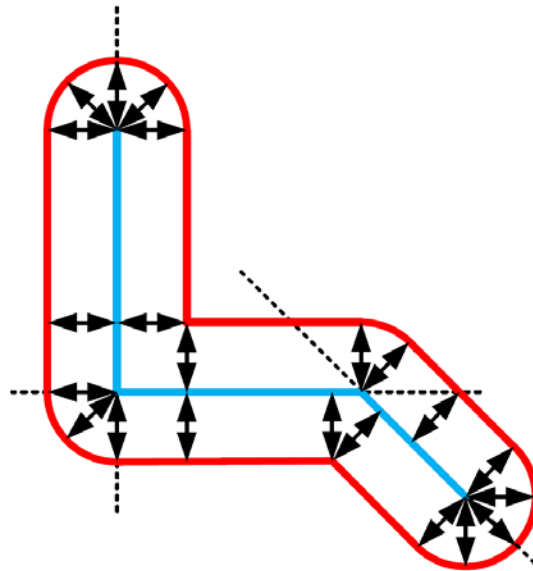
Figure 12. Illustration of equidistant deviation points (red) from a defined route (blue).

In an initial step, the algorithm selected the shortest distance between the aircraft position and each route point to derive a path deviation. Subsequently, the algorithm computed the distances between aircraft positions and route segments using two neighboring route points. This calculation determined the space between the aircraft point and an infinite line going through both route segment points. This implied that the algorithm had to assess whether the resulting perpendicular point was based on the infinite line that was part of the route segment or if it was located on one of the two adjacent line sections. If the perpendicular point was located on the defined route segment, and the great circle distance between the aircraft position and the perpendicular point was smaller than the actual shortest vector distance, the algorithm assigned a new value to the aircraft position.

To allow for distance calculations to weather areas and symbol locations, we identified all relevant coordinates for weather cell intensities and PIREP symbols. We performed this by identifying display pixel locations for weather cells and airport locations and triangulated these with display pixel locations for published airport latitude/longitude coordinates. This required scenario screenshots from the portable application for every minute of the weather scenario. The algorithm analyzed these images and determined the location of the yellow weather pixels or the brown PIREP symbol pixels. These pixel values (with a given Red, Green, Blue (RGB) value) were stored in a

vector for subsequent use in the calculation of distances to aircraft coordinates. Following the outlined procedure, the algorithm produced the shortest distances between aircraft and weather cells and the shortest distance between aircraft and PIREP symbol for each minute of the scenario.

3. RESULTS

3.1 Simulation Flights

3.1.1 Flight Profile Measures

In the present simulation, pilots flew a simulated Mooney Bravo single-engine aircraft in VMC and Marginal Visual Meteorological Conditions (MVMC) while avoiding Instrument Meteorological Conditions (IMC). At the simulation start, the flight was under VMC conditions with ceilings greater than 3,000 ft and visibility greater than 5 nmi. However, while navigating the pre-planned route, pilots encountered cloud formations and MVMC conditions with ceilings between 1,000 ft and 3,000 ft and visibility ranges from 3 nmi to 5 nmi. To avoid flying into clouds, participants had to adjust their altitude or deviate from their pre-planned route. If the participant failed to adjust the altitude or failed to make necessary deviations from the route, participants could also encounter IMC conditions with ceilings between 500 ft and 1,000 ft and visibility ranges between 1 nmi and less than 3 nmi.

To assess pilot use of portable weather applications and the effect on pilots' behavior and decision-making, we analyzed participants' vertical and horizontal flight profile. We captured the vertical flight profile by analyzing participants' altitude changes and the horizontal flight profile by analyzing participants' deviations from the pre-planned route. For both analyses, we used a Bayesian model (see Kruschke, 2014, p. 449) for a *metric-predicted* variable (i.e., altitude in feet) for two groups (experimental vs. control).

3.1.2 Altitude Changes

We used a dependent measure of altitude change by first computing a difference score between the recorded altitude for each second of the scenario with the altitude at scenario start-up (Scenario A = 6,000 ft, Scenario B = 2,000 ft). We then averaged each participant's difference scores and used one difference score per participant for the analysis.

The altitude analysis for the control group and the experimental group for Scenario A showed very similar mean altitudes, with a mean posterior altitude mode of –1,220 ft for the control group and –1,150 ft for the experimental group. On average, the control group had more altitude variations than the experimental group. However, because the posterior difference between group means is only 54.6 ft, and because the value 0 is in the center of the 95% HDI, this difference is not credible.

The altitude analysis for Scenario B also showed that both groups had very little altitude change, with a mode of –1.07 ft for the control group and a mode of 4.42 ft for the experimental group. This altitude difference is not credible, because the value 0 is included in the 95% HDI for the difference of means.

3.1.3 Deviations from the Pre-planned Route

As outlined in the Introduction, we recorded each participant's deviation from the pre-planned route every 10 seconds during the simulation. This yielded 120 deviation scores for each participant and flight scenario. We then averaged each participant's deviation scores and used one deviation score per participant for the analysis. Figure 13 illustrates the differences in the mean deviations between the control group and the experimental group.
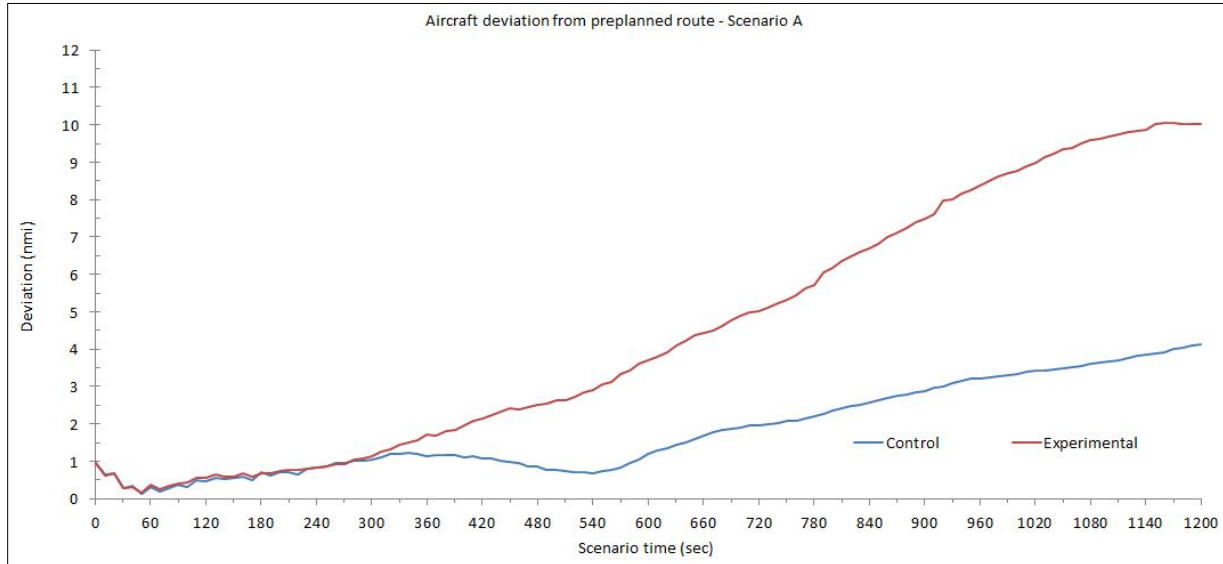
Figure 13. Mean deviation from the pre-planned route (Scenario A) for the experimental and the control groups.

Figure 14 shows the outcome of the route deviation analysis for Scenario A. As shown in the figure, the experimental group deviated more from the pre-planned route (mode = 1.82 nmi) than the control group (mode = 0.597 nmi). The difference of means (mode = 1.23) is credible, as the value 0 is not included in the 95% HDI. Furthermore, we also found the variation in deviation distances to be credibly different, with a higher SD in the experimental group (mode = 1.75) than in the control group (mode = 0.591). The posterior mean difference of the SD has a mode of 1.19, with the value 0 not included in the 95% HDI. The effect size for the differences in deviations between the groups is also credible, with a mode of 0.902 and the value 0 outside the 95% HDI.
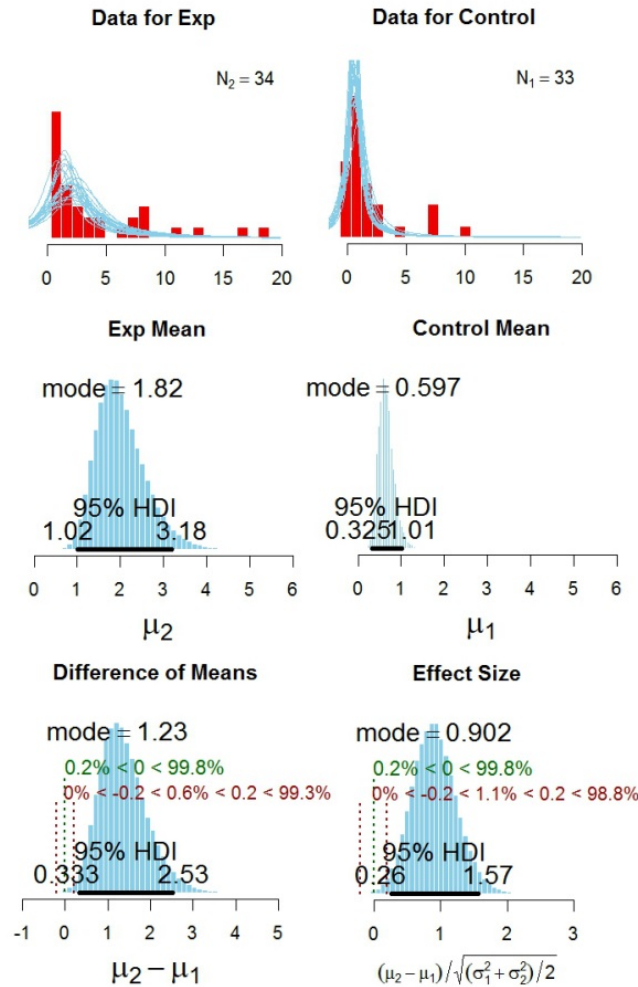
Figure 14. Scenario A data (top), posterior distributions for means (middle), difference of means (bottom left), and effect size (bottom right) for the comparison of route deviations between the experimental group and the control group.

An analysis of the route deviations for Scenario B showed the mean deviations to be similar (see Figure 15), with a mode of 1.02 nmi for the control group and 0.92 nmi for the experimental group. This difference is not credible since the value 0 is included in the 95% HDI.
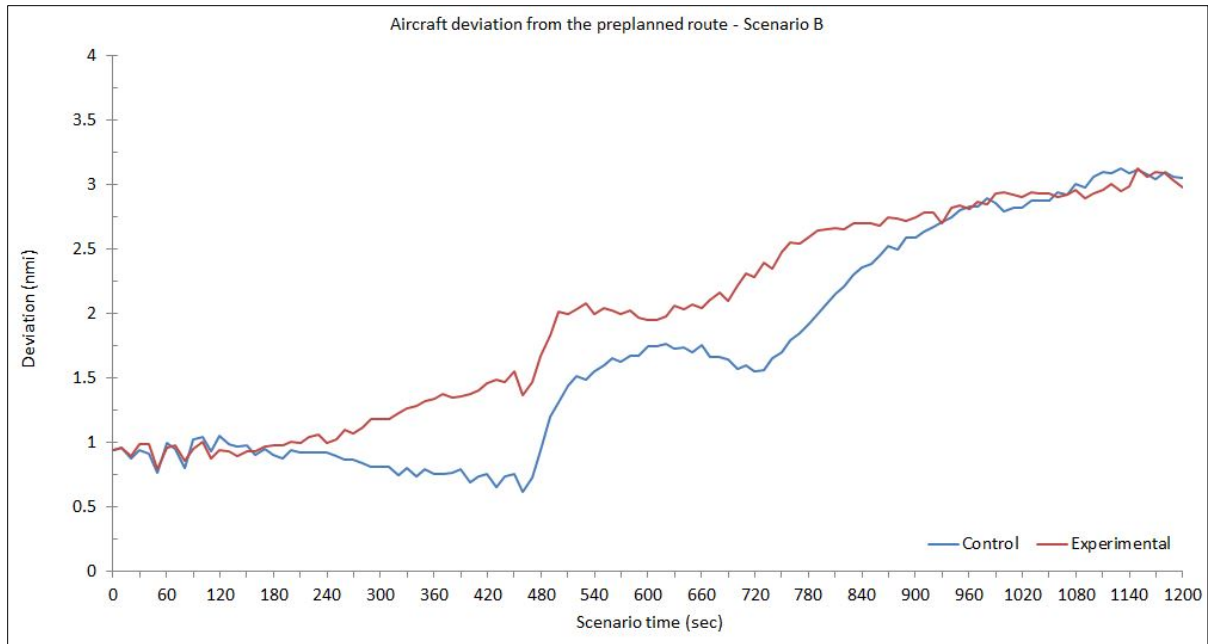
Figure 15. Mean deviation from the pre-planned route (Scenario B) for the experimental and the control groups.

When analyzing the vertical flight profiles for both scenarios, we found no credible differences in altitude changes between the experimental group (portable weather application) and the control group (no weather information). However, when analyzing the horizontal flight profiles we found a credible difference in route deviation during Scenario A. The experimental group had credibly larger deviations from the pre-planned route compared to the control group.

3.1.4 Weather Situation Awareness

One important goal of the present study is to assess how the use of portable weather applications affects pilot WSA. A high WSA would imply that a pilot is cognizant of or prepared for weather state-changes and will therefore have more time to take appropriate action. Early decision-making to avoid weather will also result in pilots keeping appropriate distances from weather events.

During the scenario flights, participants were part of a two-aircraft team ferrying an aircraft to the destination airport. They were required to communicate weather information and flight decisions to the second pilot. We logged all communications and combined each relevant message into five main categories (see Table 4).

Table 4. Communication Category Descriptors

Category				
Weather data	**Weather direct view**	**Ground view**	**Maneuver/Course change**	**Other**
- 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/decreasing 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.

Note. METAR = Meteorological Aerodrome Report; TAF = Terminal Area Forecasts; VFR = Visual Flight Rules.

The first category in Table 4, Weather data, captures all communication related to providing weather information—such as METAR, TAF, and the communication of information and weather state-changes acquired from the portable weather application. The second category, Weather direct view, captures communicated weather information acquired from the "out-the-window" view. The third category, Ground view, captures communicated information associated with terrain, landmarks, and airfields. The fourth category, Maneuver/course change, captures communications related to decisions about maneuvering the aircraft, diverting, and changing course. The last category, Other, encompasses communicated information about position, heading, altitude, intent, and other nonspecific reports.

For the following analysis, we use a combined count of transmissions from Scenario A and Scenario B. Because the analysis involves a predicted value that is a count (i.e., the number of transmissions), we used a model by Kruschke (2014, p. 703) for analysis of data on a count-valued measurement scale.

Figure 16 (left) shows the relative frequencies for the communication of Weather data information for the experimental group and the control group. With a total of $N = 137$ communications to the aircraft following for the experimental group, versus only $N = 16$ for the control group, this difference is credible (see posterior histogram on the right). Although the control group did not have access to the portable weather application, participants had some weather data information that was included in the standard weather briefing. It is this type of Weather data information that the control group participants communicated to the other aircraft.
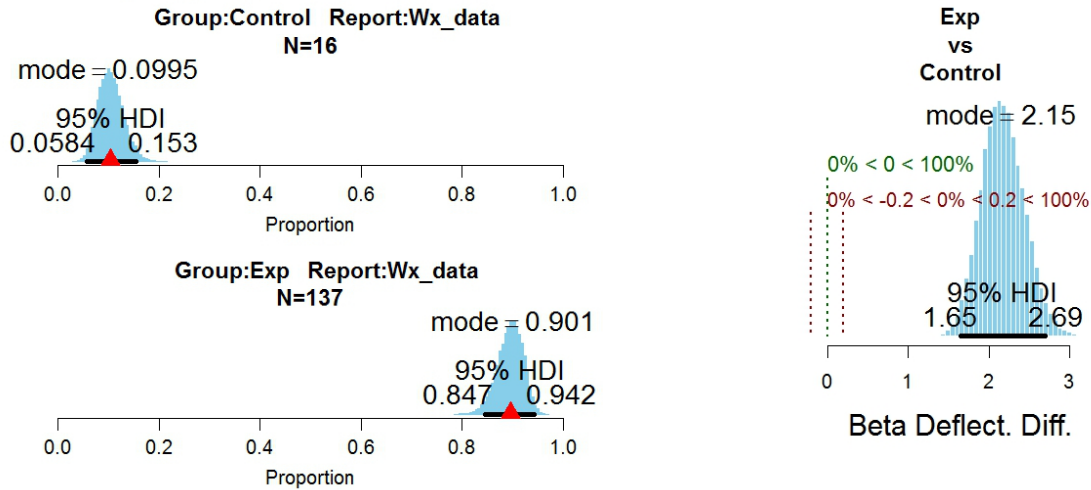
Figure 16. Posterior distributions (left) for the estimated cell proportions for the communication of **Weather data** for the experimental group and the control group. The triangles at the bottom of the histogram indicate the actual proportions for each group. The histogram to the right shows the posterior contrast for the comparison (experimental-control).

Figure 17 (left) shows the relative frequencies for the communication of Weather direct view information for the experimental group (N = 643) and the control group (N = 757). As shown in the posterior histogram (right), this difference is credible with a higher number of Weather direct view reports for the control group.
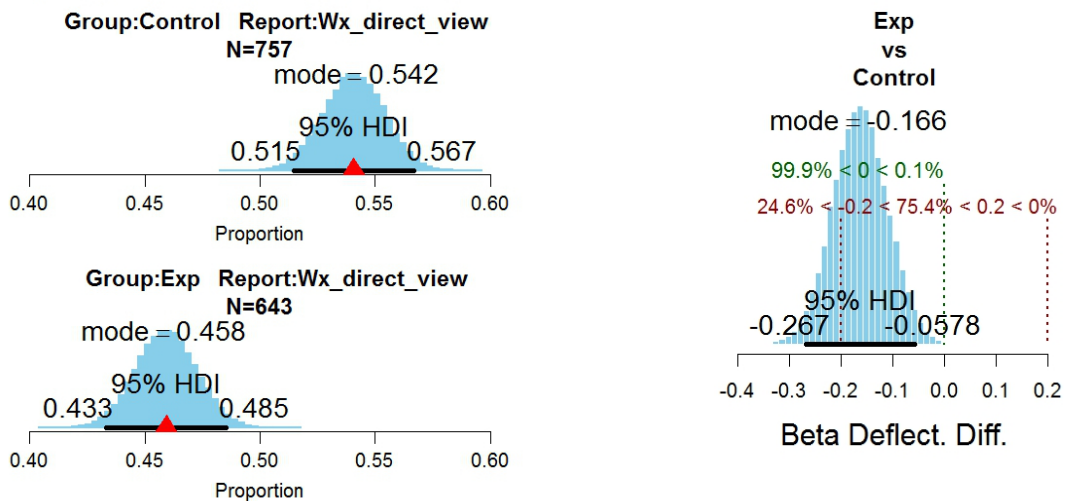
Figure 17. Posterior distributions for the estimated cell proportions for the communication of **Weather direct view** for the experimental group and the control group. The histogram to the right shows the posterior contrast for the comparison (experimental-control).

Figure 18 (left) shows the relative frequencies for the communication of Ground view information for the experimental group (N = 51) and the control group (N = 41). This difference is not credible; the value 0 is included in the 95% HDI.
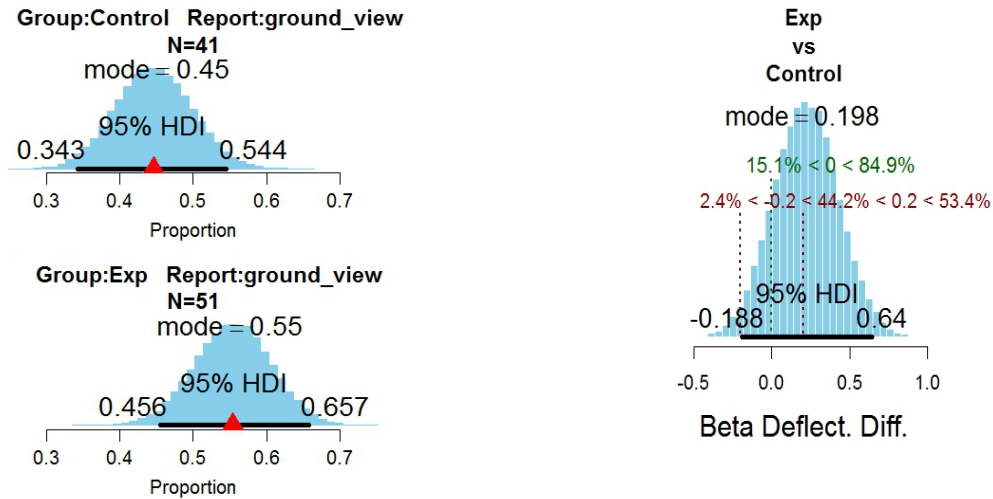
Figure 18. Posterior distributions (left) for the estimated cell proportions for the communication of **Ground view** for the experimental group and the control group. The histogram to the right shows the posterior contrast for the comparison (experimental-control).

Figure 19 (left) shows the relative frequencies for the communication of Maneuver/course change information for the experimental group (N = 306) and the control group (N = 302). This difference is not credible because the posterior histogram for the contrast (right) has the value 0 included in the 95% HDI. But even more importantly, the 95% HDI falls completely within the ROPE margins, which means that we can accept the presence of a null effect.
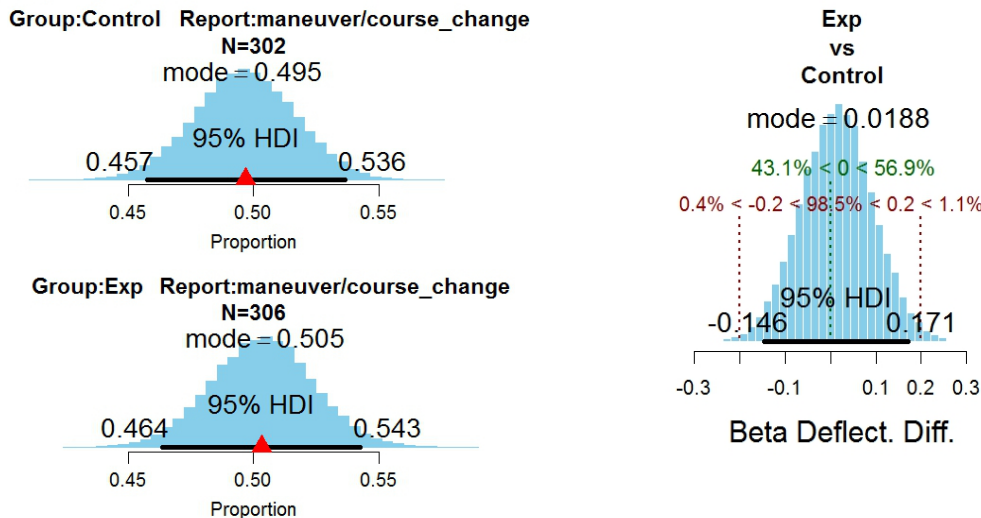
Figure 19. Posterior distributions (left) for the estimated cell proportions for the communication of **Maneuver/course change** for the experimental group and the control group. The histogram to the right shows the posterior contrast for the comparison (experimental-control).

Figure 20 (left) shows the relative frequencies for the communication of Other information for the experimental group (N = 625) and the control group (N = 633). This difference is not credible because the value 0 is included in the HDI and because the 95% HDI falls completely within the ROPE margins (i.e., a null effect).
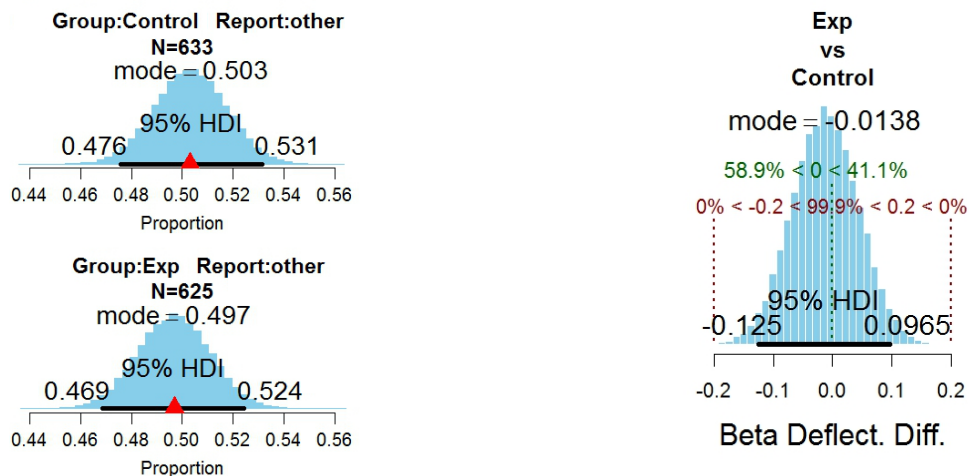
Figure 20. Posterior distributions (left) for the estimated cell proportions for the communication of **Other** for the experimental group and the control group. The histogram to the right shows the posterior contrast for the comparison (experimental-control).

The experimental group using a portable weather application provided credibly more communications of Weather data information than the control group. The control group provided a credibly higher count of communications of Weather direct view information than the experimental group. There is no credible difference in the communication of Ground view reports between the groups. For the communication of Maneuver/course change information and Other information, we find credible null effects; that is, evidence that there are no differences between groups. Taken together, this supports our hypothesis that using a portable weather application will result in an increased WSA as evidenced by the credibly larger count of communications of Weather data information for the experimental group.

3.1.5 Distance to Weather

The use of cockpit weather applications could potentially increase pilot WSA and enhance weather avoidance behavior. Current guidelines by FAA and NOAA (1983) state that hazardous weather should be avoided by at least 20 statute miles (i.e., 17.379 nmi). During a weather avoidance simulation Ahlstrom and Dworsky (2012) measured the distance-to-weather (≥ 30 dBZ precipitation cells) and found that GA pilots flew closer than the current guidelines with a mean distance-to-weather of 14 nmi.

In Scenario A, we measured the distance-to-weather (≥ 30 dBZ cells, once per minute) to assess how pilot weather avoidance behavior is affected by the use of portable weather applications (see bottom right in Figure 5 for scenario weather). In Scenario B, where no relevant precipitation hazard was present, we measured the distance from the aircraft to the center of an area where conditions were favorable for icing (as indicated on the portable application by an icing PIREP symbol). The area with the icing condition was centered, along the route flight, adjacent to an area with lower cloud ceilings and reduced visibility.

For the distance-to-icing analysis, we first averaged the 20 distance measures for each participant per flight and used one mean value per participant for the analysis. However, for Scenario A we only used the data from 10 min into the scenario until the end, as participants were too far away from the relevant precipitation area at the scenario start-up. Also, at 10 min into the scenario both the control group and the experimental group were still, on average, about 17 nmi away from the 30 dBZ cells as recommended by current guidelines (see Figure 21).
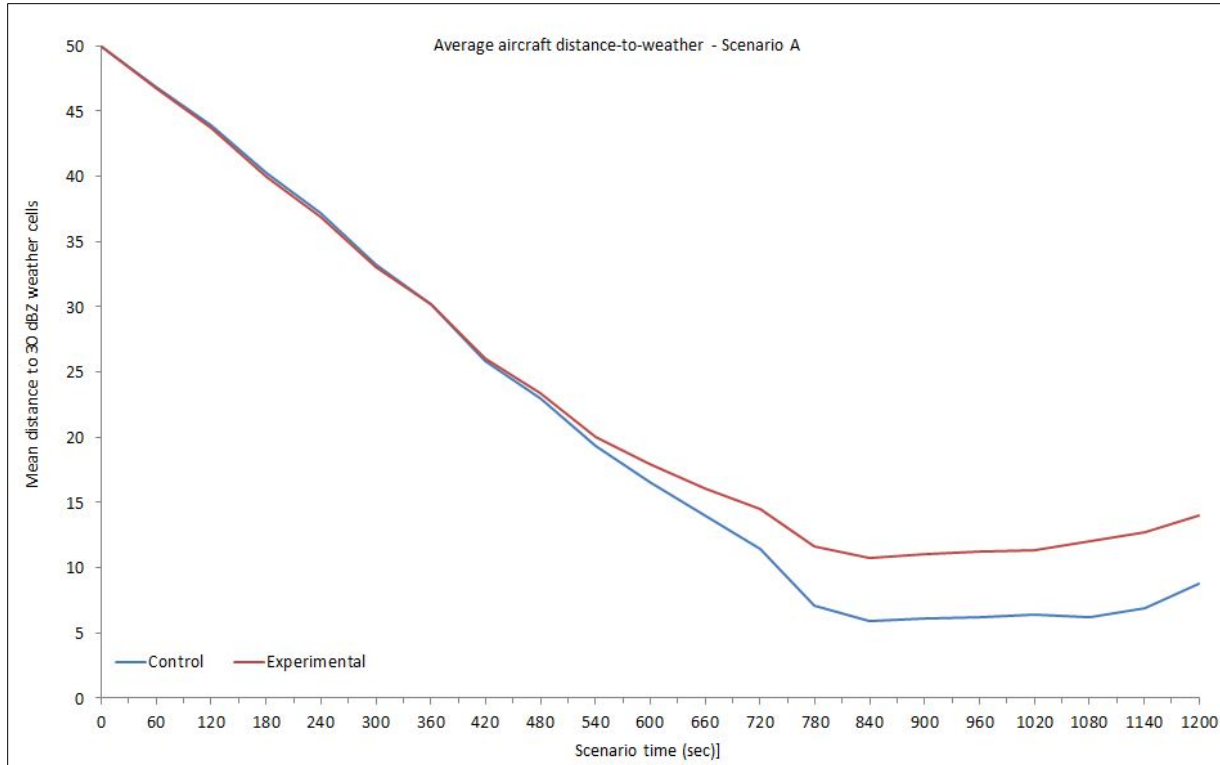
Figure 21. Illustration of the average distance-to-weather for the experimental and control group during Scenario A.

For both the Scenario A and B analyses, we used a Bayesian model (see Kruschke, 2014, p. 449) for a metric-predicted variable (i.e., distance in nmi) for two groups (experimental vs. control).

Figure 22 shows the result of the average distance-to-weather analysis for Scenario A. On average, measured for the scenario time of 10–20 min, the experimental group kept larger distances to hazardous weather cells than the control group. The posterior mean distance for the experimental group has a mode of 9.65 nmi; the posterior mode is 6.93 nmi for the control group. This difference of means is credible (mode = 2.81) as the value 0 is outside the 95% HDI. We also found a credible difference between the control group (mode = 0.758) and the experimental group (mode = 4.06) in the SD of the deviation distances, with a mean of the difference equal to 3.24. This means that while the experimental group, on average, kept larger distances away from weather than the control group, there is more variation in the deviation distances within the experimental group. Finally, there is a credible effect size with a mode of 0.911.

Figure 22. Scenario A data (top), posterior distributions for means (middle), difference of means (bottom left), and effect size (bottom right) for the comparison of the 10–40 min average scenario distance-to-weather (≥30 dBZ cells) between the experimental and the control group.

In Figure 23, we show a graphical illustration of the differences in the 10–20 min distance-to-weather. In the figure, we illustrate the average distance to the weather cell encountered at the end of Scenario A. It is important to note that the figure illustrates only the 2D proximity (i.e., distance in plan view) of each group's average flight path in relation to the pre-planned route and the weather cell. During the simulation, participants flew below the precipitation cell anvil located above their flight altitude. Many pilots commented that "It is OK and safe to fly under the storm," which is a statement that is clearly at odds with current guidelines and recommendations that state, "Don't attempt to fly under the anvil of a thunderstorm. There is a potential for severe and extreme clear air turbulence" (FAA, 2013). Nevertheless, even though the experimental group flew farther away from hazardous precipitation cells, the distance-to-weather outcome for both groups clearly shows that the participant flight behavior was far from ideal and at odds with current guidelines.
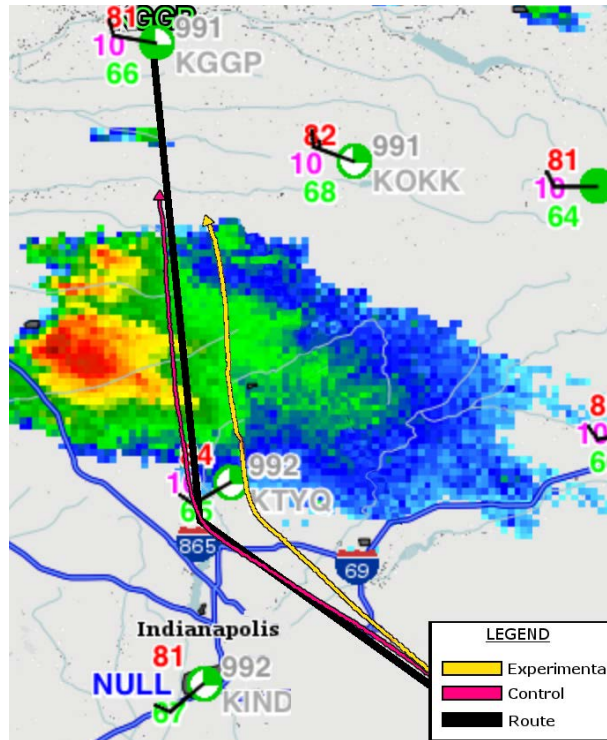
Figure 23. Graphical illustration of the average distance-to-weather for the experimental group (yellow line) and the control group (pink line) at the end of Scenario A.

From the previous analysis, we know how close participants flew to hazardous weather, on average. We also wanted to assess the closest distance that participants came to 30 dBZ cells. Figure 24 shows the outcome of the closest distance-to-weather analysis. As Figure 24 shows, the mean posterior distance for the experimental group has a mode of 5.72 nmi while the mean distance for the control group has a mode of only 2.59 nmi. The difference of means is credible with a mode of 3.06 nmi and the value 0 outside the 95% HDI. We also found a credible difference between the groups in the intra-group variability, or SD, of the closest distances (mode = 3.85). This can be seen in Figure 24 (top) where the sample data is more spread out for the experimental group (left) than the control group (right). The SD for the control group has a mode of 0.987, whereas it is 4.88 for the experimental group. Finally, there is a credible effect size for the groups' differences in the closest distance to weather with a mode of 0.869.
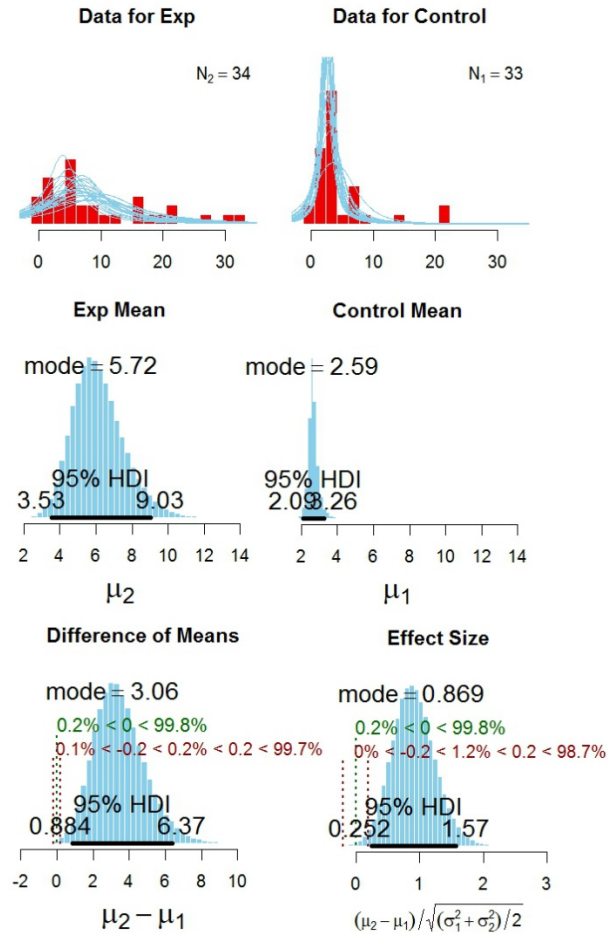
Figure 24. Scenario A data (top), posterior distributions for means (middle), difference of means (bottom left), and effect size (bottom right) for the comparison of the closest 10–40 min distance-to-weather (30 dBZ cells) between the experimental group and the control group.

These results support that the experimental group flew, on average, farther away from hazardous weather than the control group. However, both groups flew much more closely to hazardous weather than what the current FAA guidelines recommend (i.e., 17.379 nmi). Only six participants in the experimental group and two participants in the control group had the closest distances to weather that exceeded this recommendation.

3.1.6 Distance to Area of Icing

During Scenario B, low clouds and freezing levels had caused problems with icing and visibility along the pre-planned route. For the experimental group, this was indicated on their portable weather application by several PIREPs for light to moderate icing, with one PIREP symbol located directly on top of the pre-planned route (see Figure 25). At the same time, there was also a "pop-up" area adjacent to the route showing reduced visibility (see Figure 26). Although the icing PIREPs were for higher altitudes than the scenario starting altitude and the ceiling information alerted of reduced visibility from greater than 10 nmi to between 10-5 nmi, we were interested in assessing whether these information sources on the portable application would affect participant behavior in the experimental group. Because this group had weather information readily accessible, it could potentially lead to a decision to turn around or to deviate from the pre-planned route.
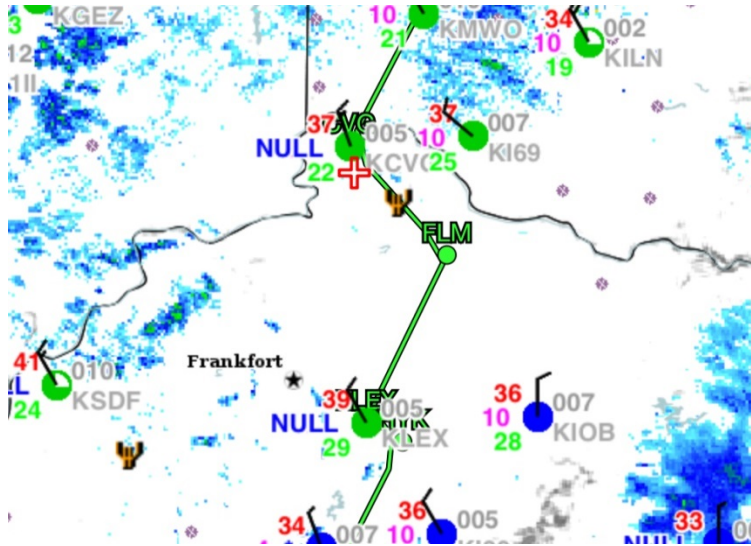
Figure 25. Illustration of the relation between the aircraft location (red cross) and the icing PIREP symbol along the pre-planned route (green line) at scenario start-up.
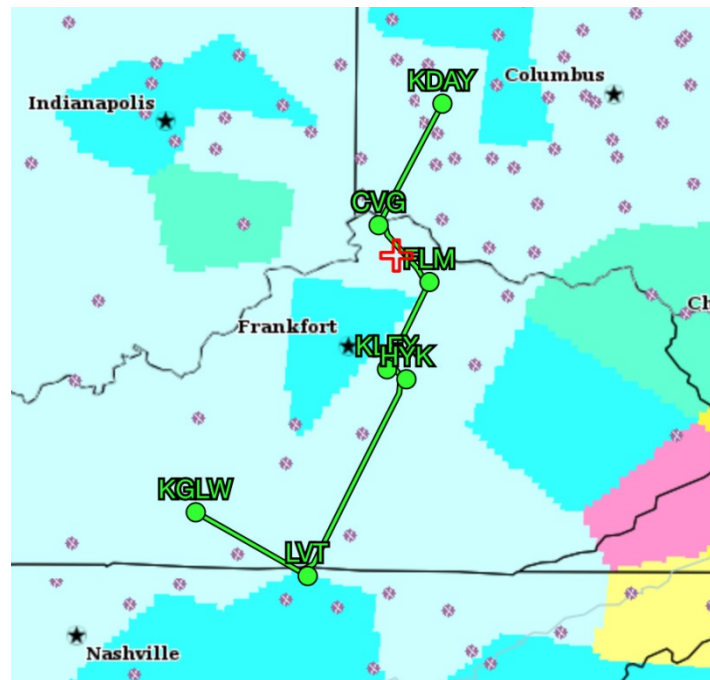
Figure 26. Illustration of the relation between the aircraft location (red cross) and the triangular "pop-up" area of reduced visibility adjacent to the pre-planned route.

First, we know from our previous analysis of Scenario B route deviations that there was no credible difference between the control group and the experimental group with regards to the average deviation from the pre-planned route. Here, we specifically assess how close pilots flew to the area centered at the icing PIREP symbol by using the closest distance for each pilot. The outcome of the shortest distance analysis shows a posterior closest distance, with a mode of 0.555 nmi

for the experimental group and a mode of 0.666 nmi for the control group. This difference is not credible because both the difference of means (mode = –0.091) and the effect size (mode = –0.258) have the value 0 included in the 95% HDIs. This means that, on average, the experimental group and the control group flew the pre-planned route without any credible deviations. Analyzing the closest distances in the sample data, we only found seven participants in the control group and eight participants in the experimental group who had greater distances than 1 nmi from the PIREP area (see Figure 27).
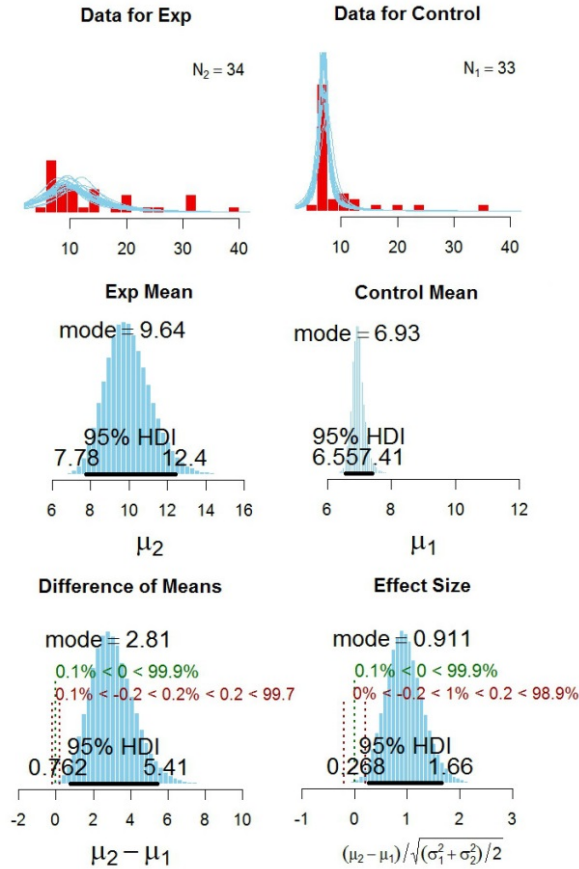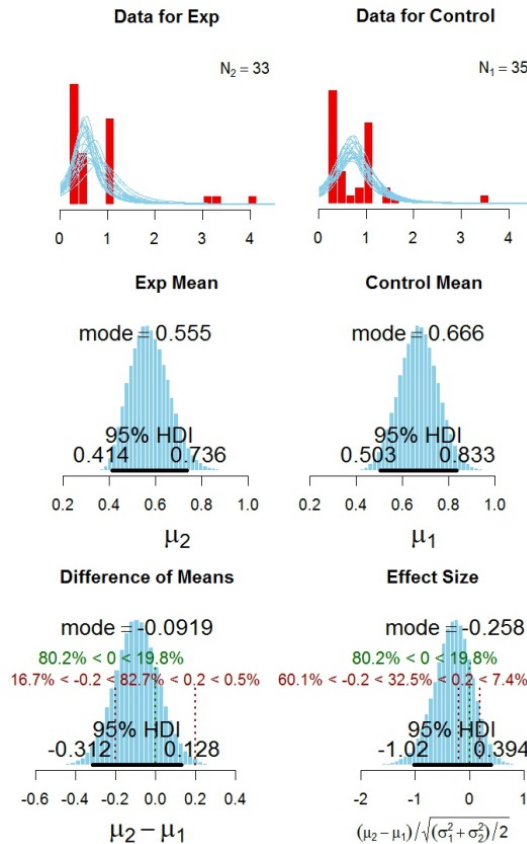

Figure 27. Scenario B data (top), posterior distributions for means (middle), difference of means (bottom left), and effect size (bottom right) for the comparison of the closest distance to the icing PIREP between the experimental group and the control group.

The results for Scenario A show there is a credible difference between groups for the average distance-to-weather where the experimental group (using a portable weather application) kept larger distances to hazardous weather cells than the control group. The control group flew credibly more closely to 30 dBZ cells (mode = 2.59 nmi) than the experimental group (mode = 5.72 nmi). Nevertheless, both groups flew too closely to hazardous weather than what is recommended in current guidelines. There was no credible difference for how closely the groups flew in relation to the icing PIREP area and the area of reduced visibility. Both groups flew the pre-planned route without any noteworthy deviations.

3.2 VFR Flight into IMC

During Scenarios A and B, participants could inadvertently enter clouds or haze where they no longer can see the horizon or the terrain. This is the dangerous VFR into IMC situation where pilots can experience spatial disorientation and lose control of the aircraft (Wiggins, Hunter, O'Hare, & Martinussen, 2012; Wilson & Sloan, 2003). In the present simulation, the VFR flights into IMC were brief and could generally be accommodated by slight changes in altitude or direction. No pilot entered IMC and lost control of the aircraft. During flights, pilots reported "Loss of VFR conditions" as they entered smaller cloud formations or entered haze in the vicinity of storm cells. There were few reports of Loss of VFR conditions for Scenario A, with only one pilot in the experimental group and one pilot in the control group. During Scenario B, there were three reports in the experimental group and one report in the control group. On average, the Loss of VFR reports came 15 minutes into the flight for both scenarios.

3.2.1 Decision-Making

One potential benefit of using a portable weather application is enhanced 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, weather locations in the near future, and the use of alternative routes to avoid hazardous weather. If a pilot is cognizant of the current weather along the route and realizes weather locations in the near future, it could shorten the time to make a decision to divert and thereby avoid hazardous weather. Therefore, we assessed the simulation time at which participants announced their decision to divert to an alternate airport due to weather. Because of the low numbers of diversion decisions, we combined the decision times for Scenario A and Scenario B. Of the 35 participants in the control group, 5 participants decided to divert. Of the 37 participants in the experimental group, 7 participants decided to divert. An analysis, using a model (Kruschke, 2014) for a metric-predicted variable (i.e., scenario time in minutes), showed mean posterior decision times for the control group and the experimental group, with modes of 12.0 minutes and 9.17 minutes, respectively. Although the experimental group made their decisions to divert earlier in the scenario, the analysis showed that the difference of means (mode = –2.95) was not credible, because the value 0 was included in the 95% HDI (extending from –14.6 to 8.54).

3.2.2 Weather Presentation Interaction

During the simulation, we recorded all interactions with the portable weather application. The application allows pilots to overlay weather information from one of the following layers: flight categories (e.g., VFR, IFR), ceiling information, visibility, precipitation, icing probability, turbulence potential, wind, temperature, relative humidity, and satellite imagery. Figure 28 shows the proportion of time spent in each of the layers for Scenario A (convective storm encounter) versus Scenario B (icing, ceiling, and visibility).
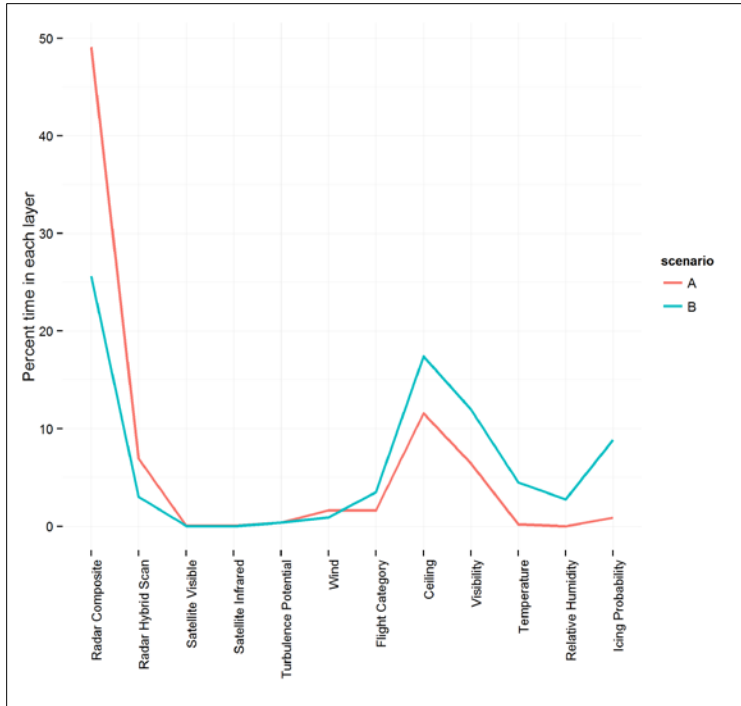
Figure 28. Percentage of scenario time spent using each of the 12 possible layers.

From the display times shown in Figure 28, it is clear that participants mostly used the same information during Scenarios A and B. The majority of the display time during Scenario A (convective scenario) is for precipitation information, which accounts for roughly half of the total display time. Besides precipitation, pilots displayed ceiling information and visibility information for roughly 12% and 6% of the total display time, respectively. The display time was minimal for weather information sources such as wind, flight category, temperature, relative humidity, and icing probability. For Scenario B (ceiling, icing, and visibility scenario), the precipitation information only accounts for roughly a quarter of the total display time. Instead, compared to Scenario A, pilots had longer display times for information of flight category, ceiling, visibility, temperature, relative humidity, and icing probability.

In addition to a single background layer, pilots were able to independently toggle three additional features on or off in the portable application. These features consisted of symbols indicating METAR information, TAFs, and PIREPs. The total duration over which each feature was visible was analyzed using a model (Kruschke, 2014, p. 583) for a metric-predicted variable (i.e., visible duration in minutes) and two factors (i.e., feature type and scenario). There was no difference in feature usage during the two scenarios (the mode of the posterior distribution of the difference in means was –4.6, and the 95% HDI included 0), nor was there a credible interaction between feature and scenario (the mode of the posterior distribution was –0.3).

3.2.3 Cognitive Engagement

Cognitive engagement was measured using fNIR, which was sampled at 16 channels located on the forehead at a rate of 2 Hz. Changes in the volume of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (Hb) are inferred based on the absorption of two frequencies of infrared light. Figure 29 shows the time course of the mean level of oxygenated blood (averaged across channels and participants) for pilots in the experimental and control groups in Scenario A. Pilots using the portable weather application showed a higher level of cognitive engagement, evidenced by a higher mean blood oxygenation level, for the duration of the scenario.
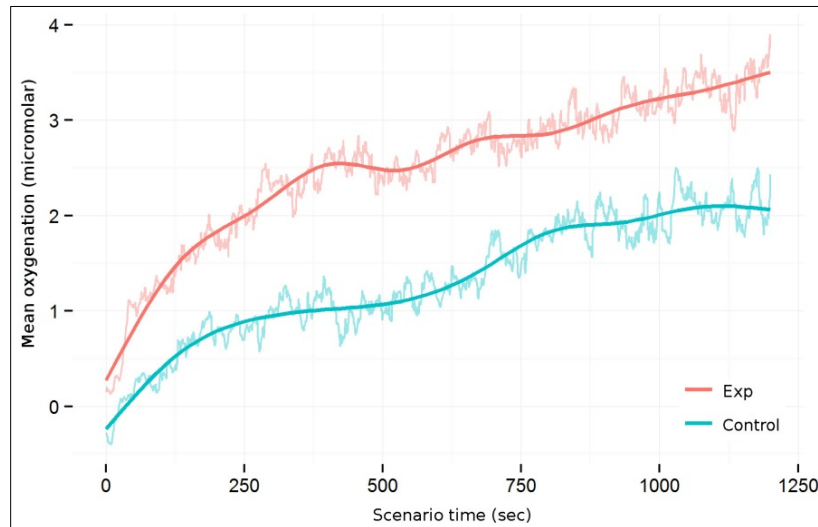
Figure 29. Scenario A - fNIR data for the experimental group and the control group.

We used a Bayesian model (see Kruschke, 2014, p. 449) for a metric-predicted variable (i.e., oxygenation values) for two groups (experimental vs. control) to examine differences in the mean oxygenation level across the entire scenario. As Figure 30 shows, the mean posterior oxygenation level for the experimental group has a mode of 2.5 µmol/L, whereas the mean level for the control group has a mode of 1.4 µmol/L. The difference of means is credible, with a mode of 1.1 and the value 0 outside the 95% HDI. There is also a credible effect size, with a mode of 0.706 and the value 0 outside the 95% HDI.
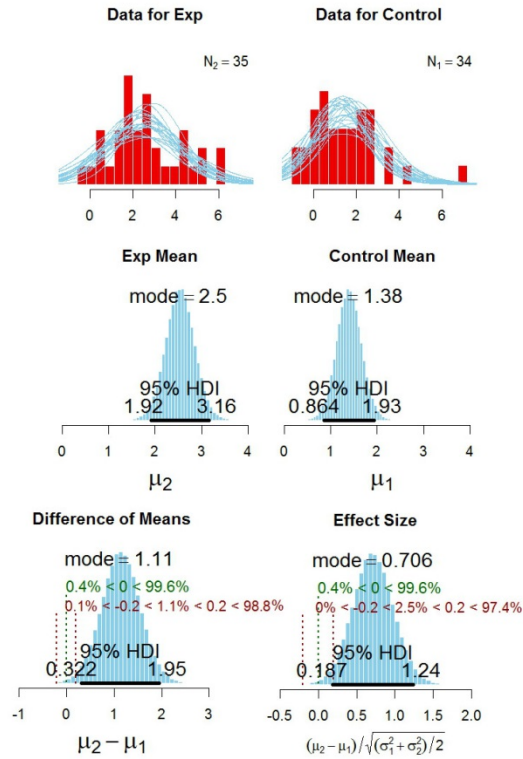
Figure 30. Scenario A - fNIR data (top), posterior distributions for means (middle), difference of means (bottom left), and effect size (bottom right) for the comparison of oxygenation changes between the experimental group and the control group.

Figure 31 shows the time course of the mean level of oxygenated blood (averaged across channels and participants) for pilots in the experimental and control groups during Scenario B. However, in this scenario, the oxygenation difference between the experiment group and control group is not as large compared to the oxygenation difference between the experimental group and the control group in Scenario A.
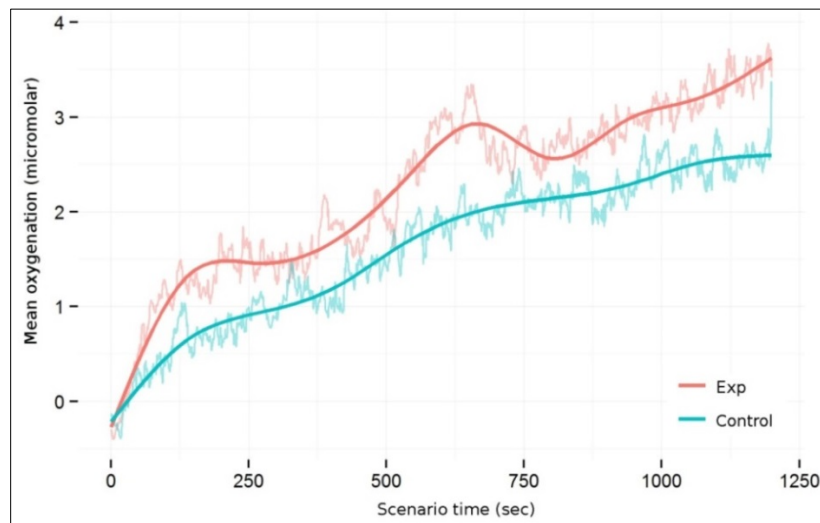
Figure 31. Scenario B - fNIR data for the experimental group and the control group.

This was confirmed following a Bayesian estimation of the mean oxygenation level for the duration of the scenario. Figure 32 shows the posterior distribution of the mean oxygenation level for the experimental group (with a mode of 2.1 μmol/L) and the mean oxygenation level for the control group (with a mode of 1.5 μmol/L). This difference is not credible because both the difference of means (mode = 0.67) and the effect size (mode = 0.35) have the value 0 included in the 95% HDIs. Therefore, the average oxygenation levels for participants in the experimental group and the control group were not credibly different.
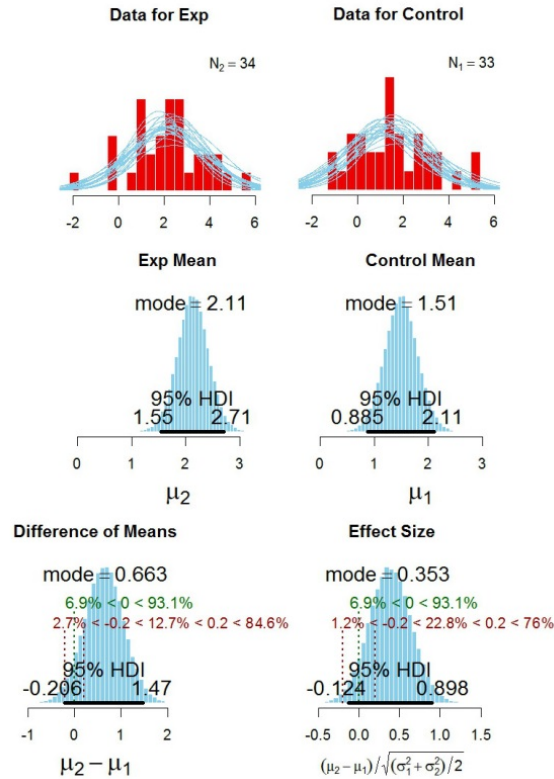
Figure 32. Scenario B fNIR data (top), posterior distributions for means (middle), difference of means (bottom left), and effect size (bottom right) for the comparison of oxygenation changes between the experimental group and the control group.

3.2.4 Post-Scenario Questionnaire

All participants completed a 10-item rating-scale questionnaire (see Appendix C); an additional yes or no item plus an open-answer item was provided to report any discomfort with the fNIR sensor. These results are presented in Appendix C. For the analysis of the ratings, we used a model by Kruschke (2014, p. 682) for an ordinal predicted variable (i.e., questionnaire ratings) comparing two groups. No credible differences between control-group and experimental-group pilots were seen in the frequency or ease of use, the effectiveness of weather information, the trust in the displayed information, or the reported mental workload.

3.3 Change-Detection of Weather Elements

The change-detection experiment assessed pilot sensitivity to changes in weather elements from the portable weather application. We used images for cloud ceiling, precipitation, and PIREP information. Because a factorial combination of all the weather elements was beyond the scope of

this study, we chose only three weather elements for the change-detection assessment. An illustration of a PIREP signal trial is illustrated in Figure 33. The left side shows a plain background map with the route (Image 1) and the right side shows the same map background with added PIREP information for turbulence and icing (Image 2).
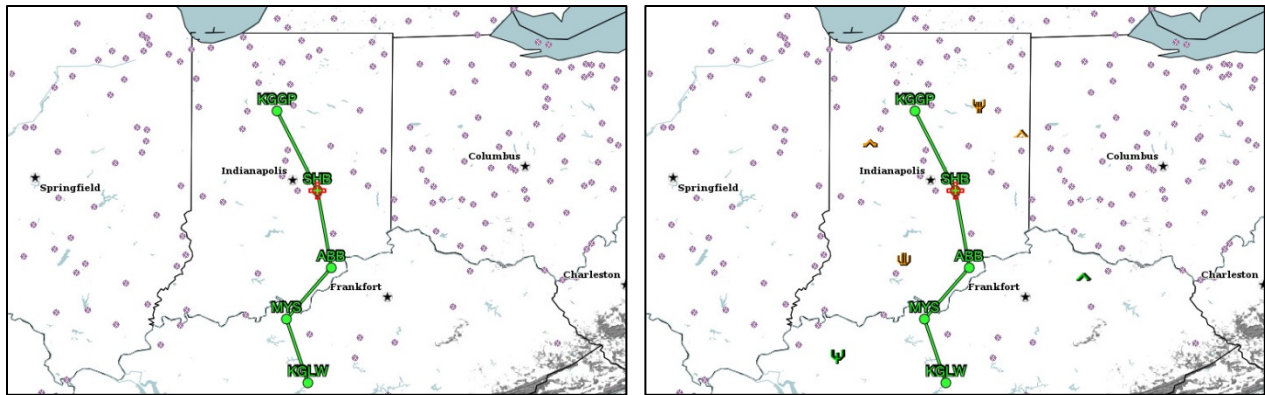
Figure 33. Illustration of a PIREP signal trial using the portable application map background (Image 1 to the left and Image 2 to the right). The PIREPs appear in Image 2 – defined as a stimulus onset trial.

For the experiment, we created 40 unique trials by pairing specific weather images with different visible features and with different levels of weather activity. The experiment consisted of 40 signal trials in which a change occurred and 40 noise trials with no change. Half of the signal trials featured a symbol onset (the signal appeared in Image 2), while the remaining trials featured a symbol offset (the signal appeared in Image 1). During the experiment, each participant viewed and responded to each unique trial two times, yielding a total of 80 trials. The trials were presented in random order. The signal trials with feature onset changes are described in Table 5.

Table 5. Signal Trials for Experiment 1

Weather element change	Trial type	Image 1	Image 2
Precipitation appearance	Signal-onset	Route display with no weather	Image with 100% precipitation cells
Precipitation appearance	Signal-onset	Route display with no weather	Image with 50% precipitation cells
Cloud ceiling color change appearance	Signal-onset	Route display with no weather	Cloud ceiling with 100% low-ceiling areas
Cloud ceiling color change appearance	Signal-onset	Route display with no weather	Cloud ceiling with 50% low-ceiling areas
PIREP appearance with no precipitation	Signal-onset	Route display with no weather	Route display with 6 PIREPS
PIREP appearance with no precipitation	Signal-onset	Route display with no weather	Route display with 3 PIREPS
PIREP appearance over precipitation	Signal-onset	Image with 100% precipitation cells	PIREPS and 100% precipitation cells
PIREP appearance over precipitation	Signal-onset	Image with 50% precipitation cells	PIREPS and 50% precipitation cells
PIREP intensity change over precipitation	Signal-onset	PIREPs and 100% precipitation cells	PIREPs intensity change and 100% precipitation cells
PIREP intensity change over precipitation	Signal-onset	PIREPs and 50% precipitation cells	PIREPs intensity change and 50% precipitation cells

For the experiment, we labeled all change trials as signal trials (Table 5). We also included an equal number of trials where there was no change between Image 1 and Image 2. We labeled these trials as noise trials. The noise trials were created by displaying the same signal image for both Image 1 and Image 2. Because we are using YES and NO responses during the change-detection

experiment, we labeled each response as a hit if the participant responded YES to a signal trial, a false alarm if the participant responded YES to a noise trial, a miss if the participant responded NO to a signal trial, and correct rejection if the participant responded NO to a noise trial.

From the recorded counts of hits, false alarms, misses, and correct rejections, we derived indices of discriminability (d) and bias (c) using a Bayesian Signal Detection Theory (SDT) model from Lee (2008). The discriminability index, d, measures how easily participants can distinguish signal trials (change) and noise trials (no change). The higher the d value is, the easier it was for participants to detect weather element changes. The bias index, c, is a measure of the participant decision-making criterion. If a participant has a positive value of the bias index, the participant has a bias to respond NO. This will result in an increase in the number of correct rejections, but it will also increase the number of misses. If a participant has a negative bias index, the participant displays a bias towards answering YES, and this leads to an increase in the number of hits but also an increase in the number of false alarms.

Even though the d and the c indices are meaningful from a psychological performance standpoint, it can nevertheless be difficult to grasp exactly what these indices imply for participant performance. For example, if a group of participants have a posterior average of $d = 1.8$ and $c = 0.8$, we can infer that the performance for this group is worse compared to another group that has a posterior average of $d = 2.8$ and $c = 0.1$. But how good is $d = 2.8$? And how bad is $d = 1.8$?

To help the reader interpret the outcomes, we introduce the concept of *ideal observers* (Olman & Kersten, 2004). An ideal observer is a hypothetical concept where we imagine a human observer that performs optimally on a task given all the available information. In the present context, it means an observer with ideal response characteristics—that is, an observer that has no response bias and that always discriminates between signal and noise images with perfect accuracy. For each change-detection condition, we derived the d and the c indices for a modelled group of ideal observers (using the same N and total number of signal and noise trials as were administered to the study participants). In each result graph, we indicate the perfect performance for the group of ideal observers so that readers can gauge the performance of the study participants. The analysis for each change-detection condition shows the posterior d and c derived from participant performances. The indication for the ideal observers, therefore, shows the best performance possible when study participants performed with zero bias and perfect discrimination. There were a few participants that came close to optimal performance—but the performance of the majority of participants were far off from this benchmark.

In addition to analyzing discriminability and bias, we also analyzed the response time for each change-detection trial during the experiment. In general, well-designed symbols and weather elements that produce good legibility and salience against the background are easier to detect and require less time to discriminate than elements with less optimal stimulus and background characteristics (McDougall, de Bruijn, & Curry, 2000). For the response time analyses, we used the combined response times for each signal trial and noise trial for each display element and participant.

3.3.1 Cloud Ceiling Color Change

Several layers of the portable weather application use colored areas to differentiate dissimilar regions. For example, the Flight Category layer separates areas of VFR, Marginal Visual Flight Rules (MVFR), IFR, and Low Instrument Flight Rules (LIFR) by solid shading of blue, green, yellow, and pink, respectively. Similarly, the Ceiling layer separates areas with different ceilings using the same color-shading principle. Because we have no prior data on participant sensitivity to changes of color-shaded areas, we assessed participant discriminability of signal and noise trials in two Ceiling layer conditions. To create two different ceiling image conditions that varied in the number of color-

shaded areas, we first used an unaltered Ceiling layer image (100%) with color-shaded areas representing visibilities ranging from unlimited (light blue) to 200 ft (pink). To create a second image condition with fewer color-shaded areas (50%), we reduced the unaltered image to include only half of the areas by omitting color-shading for ceiling levels ranging between 5,000–10,000 ft and 2,000–3,000 ft (see Figure 34).
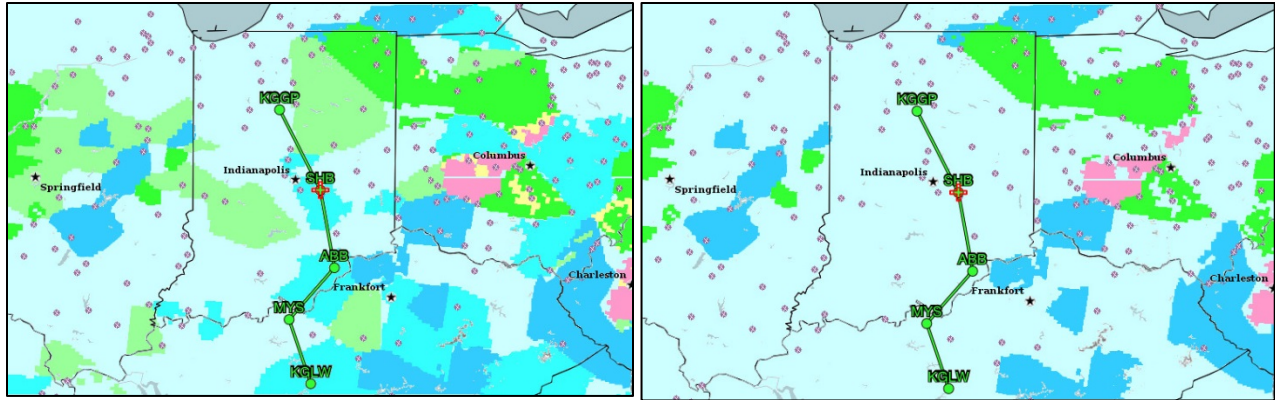
Figure 34. Illustration of signal images for the unaltered (100%, left) and the reduced (50%, right) ceiling areas plotted on the light blue display map background.

Figure 35 shows a summary of the SDT analysis in terms of discriminability, bias, hits, and false alarms. Although the hit rate is high, there are false alarms, which are taken into account when deriving the discriminability index, *d*.

Figure 35. Analysis summary with discriminability (*d*), bias (*c*), hit rate, and false-alarm rate for the comparison of 100% (red line) and 50% (blue dashed line) Ceiling layer changes.

The analysis revealed moderate discriminability and low negative bias for both the 100% and 50% trials (see Figure 36), with a mean posterior d of 2.85 ($c = -0.214$) and 2.92 ($c = -0.15$) for the 100% and 50% conditions, respectively. Nevertheless, the participants' discrimination performance is modest in comparison to the group of ideal observers. Because the posterior mean difference between the 100% and the 50% conditions had a mean of -0.0634 and the 95% HDI included the value 0, there is no credible difference in discriminability between the 100% (unaltered) and 50% (reduced) ceiling conditions.
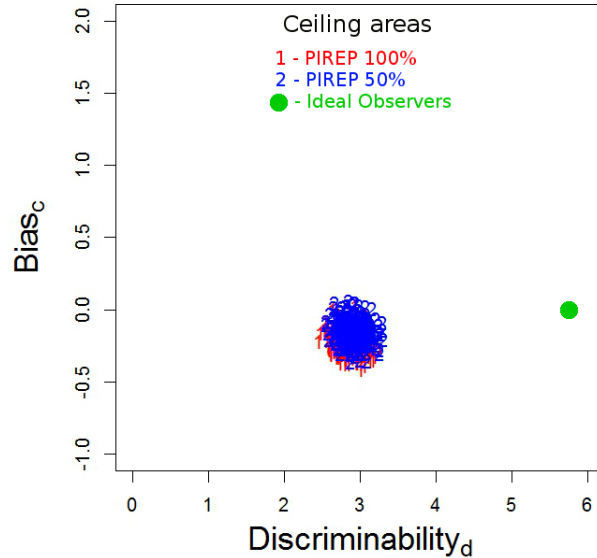
Figure 36. Discriminability (*d*) and bias (*c*) for a comparison of the 100% and 50% Ceiling area conditions. The green dot indicates the mean posterior *d* and *c* for a group of ideal observers.

The average response time for the 100% and 50% Ceiling conditions were also similar, with sample median response times of 1.451 seconds and 1.513 seconds, respectively. Before comparing the 100% and 50% response times, we performed a $\log_{10}$ transformation on the skewed sample distributions. The analysis revealed that the posterior difference distribution of the log transformed response times was not credibly different from 0 (mode = -0.015). This means that the participants' response times were equivalent for the two Ceiling conditions.

3.3.2 Number of PIREP Symbols

The present study introduced a weather presentation symbol not assessed in the Ahlstrom and Dworsky (2012) and Ahlstrom and Suss (2014) studies: PIREP symbols for reported encounters with icing and turbulence conditions. Here, we assessed participant sensitivity to changes in the (a) number of PIREP symbols presented without precipitation information, (b) PIREP symbols presented along with two levels of precipitation information, and (c) PIREP symbol intensity changes for two levels of precipitation information.

In the first condition, we manipulated the number of PIREP symbols presented along with a route segment on the map background (see Figure 37). In a high number condition, we used six PIREP symbols in the image (called the 100% image). To create a low number condition, we reduced the six PIREP symbols image to include only three PIREP symbols (called the 50% image). The left image in Figure 37 illustrates the condition with six PIREP symbols (100%), and the right image illustrates the condition with three PIREP symbols (50%).
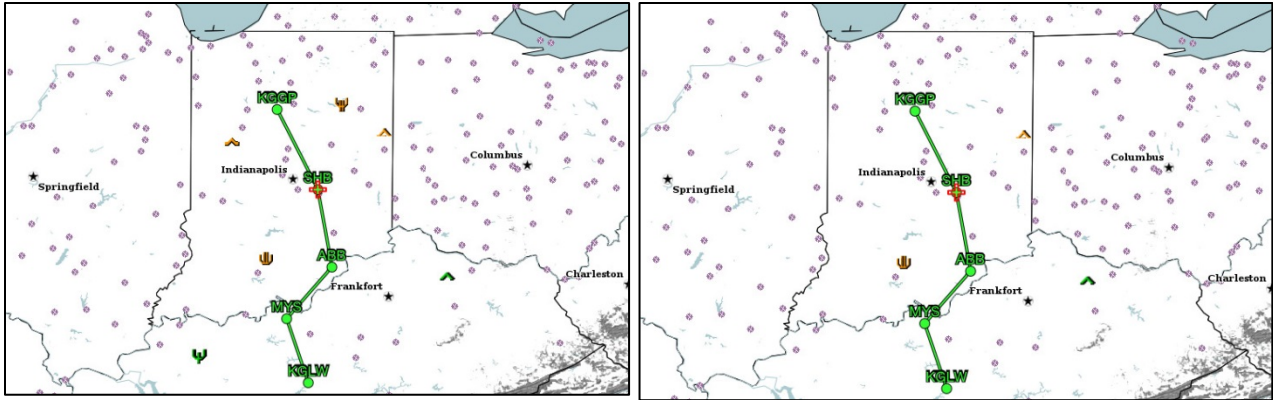
Figure 37. Illustration of signal images for 100% (left) and 50% (right) PIREP symbols plotted with a route segment on the display map background.

Figure 38 shows a summary of the SDT analysis in terms of discriminability, bias, hits, and false alarms.

Figure 38. Analysis summary with discriminability (*d*), bias (*c*), hit rate, and false-alarm rate for the comparison of 100% (red line) and 50% (blue dashed line) PIREP symbol changes.

As Figure 39 shows, an analysis of the data revealed a positive bias and an equivalent discriminability for the 100% and 50% PIREP conditions (mean posterior *d* = 1.96 for both conditions). Compared to the group of ideal observers, the participant discrimination performance is very low. The average response time for the 100% and 50% PIREP conditions was also similar, with sample median response times of 1.638 seconds and 1.622 seconds, respectively. The posterior difference distribution of the log transformed response times was not credibly different from 0 (mode = –0.0001), implying that participants' response times were equivalent for the two PIREP conditions.
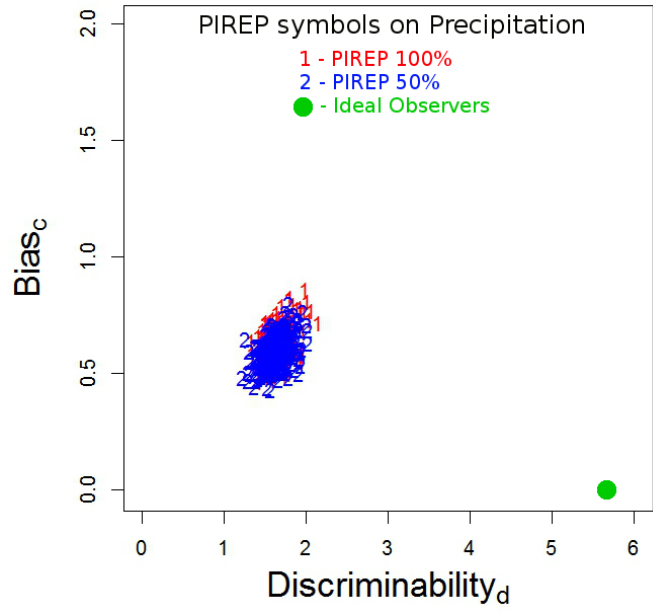
Figure 39. Discriminability (*d*) and bias (*c*) for a comparison of the 100% (6 symbols) and 50% (3 symbols) PIREP conditions. The green dot indicates the mean posterior *d* and *c* for a group of ideal observers.

3.3.3 PIREP Symbols and Precipitation Information

In the second PIREP condition, we presented six PIREP symbols along with two levels of Precipitation (see Figure 40). One PIREP image contained a large number of precipitation cells (called the 100% Precipitation), while the other type of PIREP image contained only half the number of precipitation cells of the 100% image (called the 50% Precipitation). The left image in Figure 40 illustrates the condition with six PIREP symbols and 100% Precipitation. The right image illustrates the condition also with six PIREP symbols but with only 50% Precipitation.
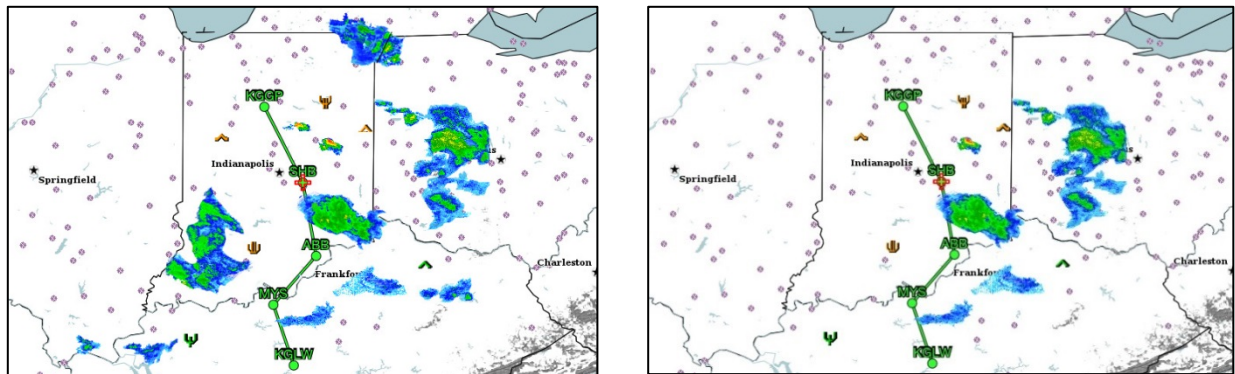
Figure 40. Illustration of signal images for six PIREP symbols and 100% Precipitation (left) and six PIREP symbols and 50% Precipitation (right).

40

Figure 41 shows a summary of the SDT analysis in terms of discriminability, bias, hits, and false alarms.
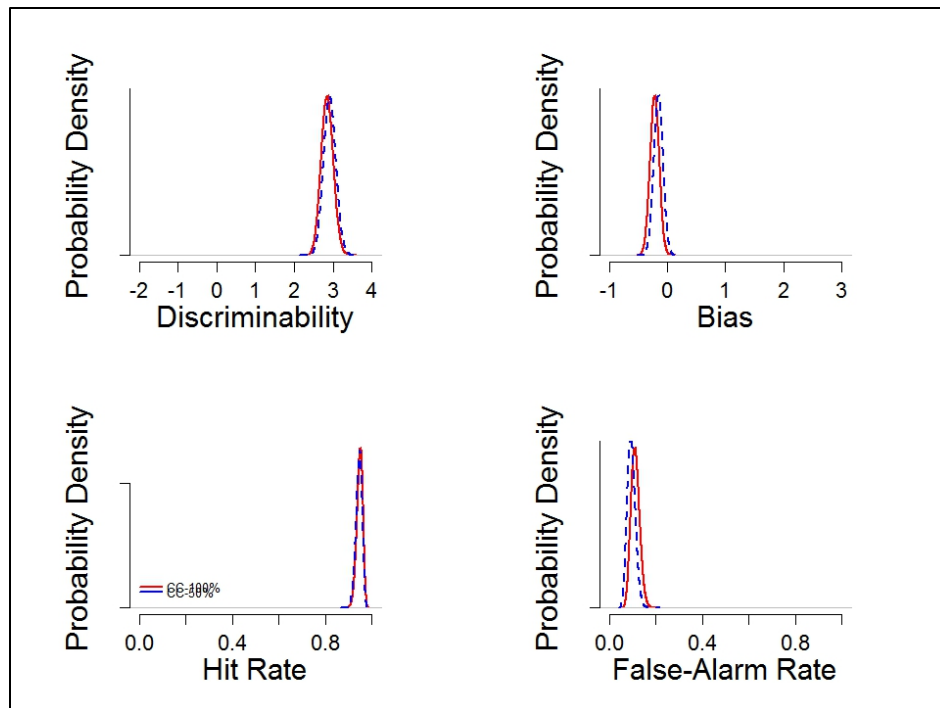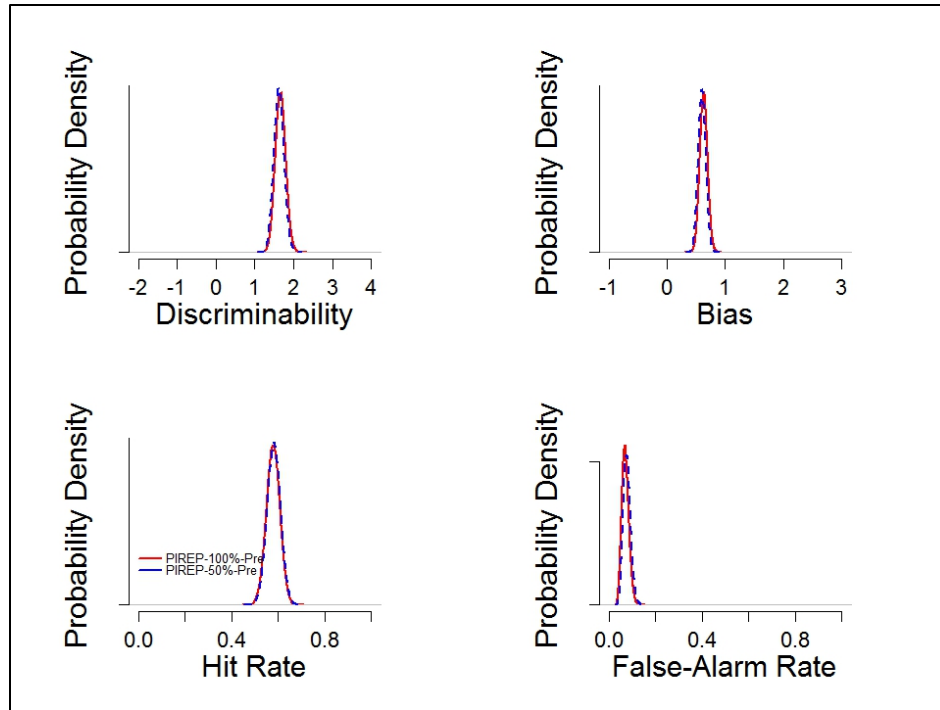

Figure 41. Analysis summary with discriminability (*d*), bias (*c*), hit rate, and false-alarm rate for the comparison of 100% (red line) and 50% (blue dashed line) precipitation and PIREP symbol changes.

An analysis revealed a positive bias and an equivalent discriminability for PIREP symbols on the 100% and 50% Precipitation cells, with mean posterior $d = 1.67$ and $d = 1.63$, respectively (see Figure 42). Again, compared to the performance of the group of ideal observers the participant performance is poor. The average response time for the PIREP change detection on the 100% and 50% Precipitation cells was also similar, with sample median response times of 1.731 seconds and 1.669 seconds, respectively. The posterior difference distribution of the log transformed response times was not credibly different from 0 (mode = −0.012), implying that participants' response times were equivalent for the two PIREP conditions on Precipitation.
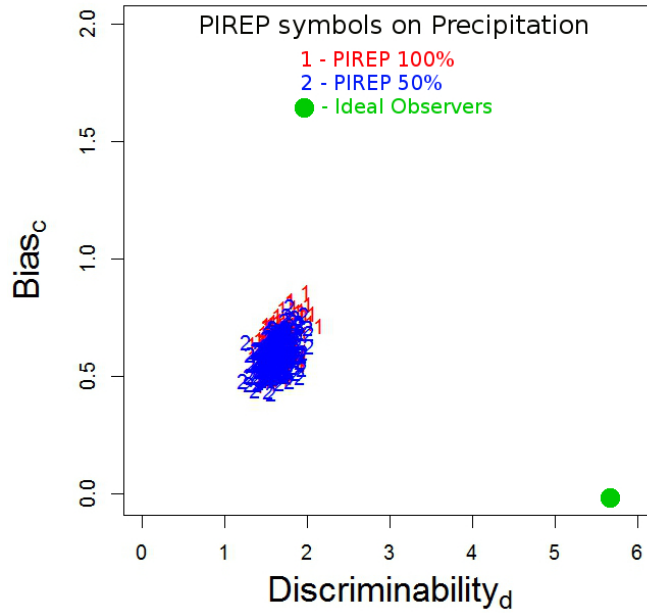
Figure 42. Discriminability (*d*) and bias (*c*) for a comparison of 6 PIREP symbols on 100% and 50% Precipitation. The green dot indicates the mean posterior *d* and *c* for a group of ideal observers).

3.3.4 PIREP Symbol Intensity Change and Precipitation

In the last two PIREP conditions, we assessed participant sensitivity to icing and turbulence PIREP symbols that changed intensity between Image 1 and Image 2 while being displayed along with the same 100% and 50% Precipitation information used in the previous PIREP conditions. For example, in Image 1 a given PIREP symbol might indicate light icing, but in Image 2 the same PIREP symbol changed to indicate moderate icing (as illustrated in Figure 43).
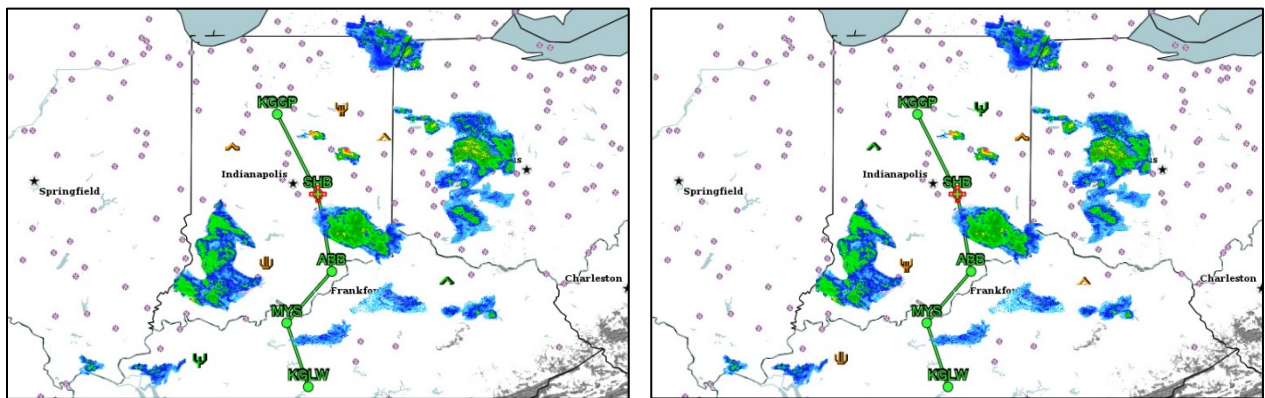
Figure 43. Illustration of signal images for icing and turbulence PIREP symbols on 100% Precipitation that change intensity, but not position, from Image1 (left) to Image 2 (right). The second condition used the same PIREP configuration but with only 50% precipitation.

42

Figure 44 shows a summary of the SDT analysis in terms of discriminability, bias, hits, and false alarms for the PIREP intensity change.
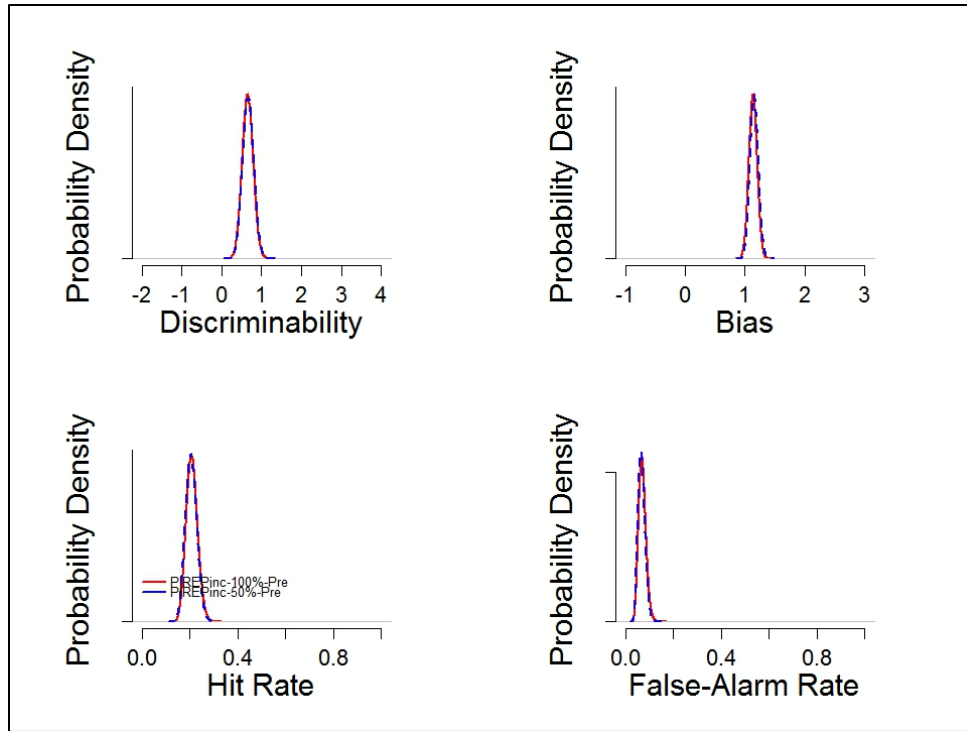
Figure 44. Analysis summary with discriminability (*d*), bias (*c*), hit rate, and false-alarm rate for the comparison of 100% (red line) and 50% (blue dashed line) precipitation and PIREP intensity symbol changes.

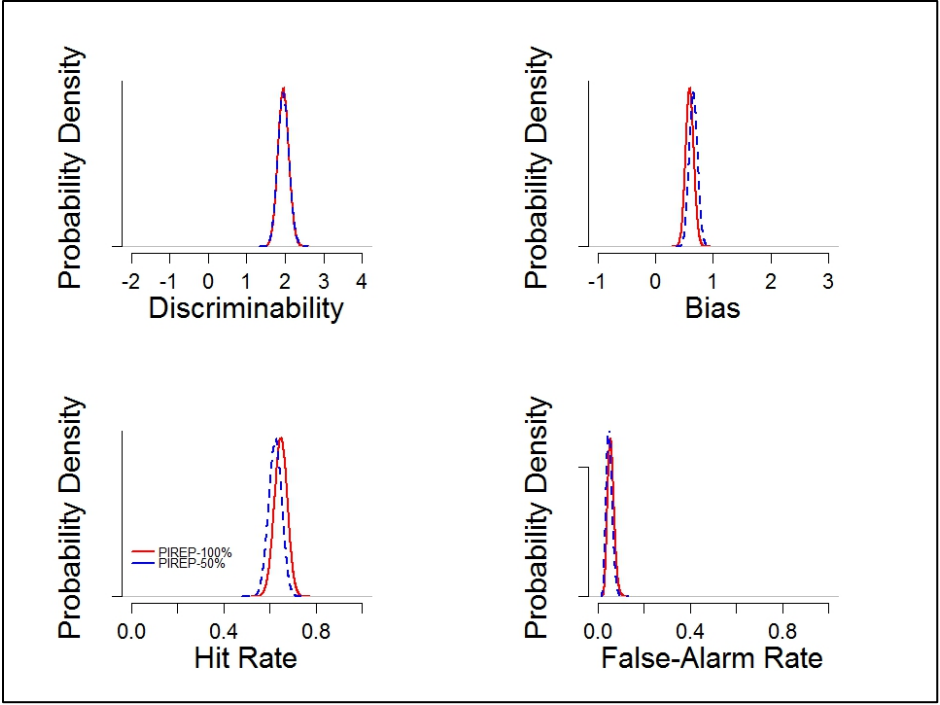An analysis revealed a positive bias and an equivalent discriminability for PIREP symbols on the 100% and 50% Precipitation cells, mean posterior were $c = 1.14$ and $d = 0.659$, and $c = 1.16$ and $d = 0.673$, respectively (see Figure 45). The average response time for the PIREP intensity change on the 100% and 50% Precipitation cells had median sample response times of 1.903 and 1.763 sec, respectively. Compared to the performance of the ideal observers, the participant performance is at the opposite end of the performance scale. The posterior difference distribution of the log transformed response times showed a credible difference between the 100% and 50% precipitation conditions (mode = –0.0186), with longer response times for the 100% precipitation condition.
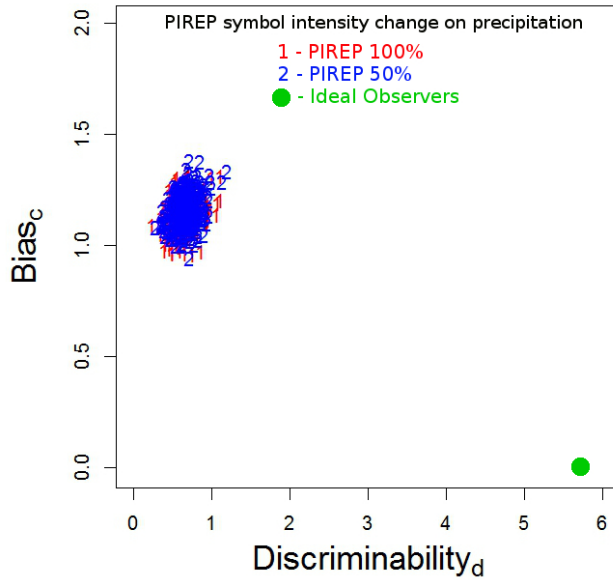
Figure 45. Discriminability (*d*) and bias (*c*) for a comparison of PIREP symbols that changed intensity on 100% and 50% Precipitation. The green dot indicates the mean posterior *d* and *c* for a group of ideal observers.

3.4 Summary of Study Findings

We designed the study to support or refute our hypotheses about the effect of portable weather presentations on WSA and flight behavior and to evaluate pilot sensitivity to weather element changes. In this section, we summarize the support, or lack of support, for each simulation hypothesis as well as illustrate this analysis in a simple tabular format in Table 6. In addition, we summarize the outcome of the change-detection experiment and the outcome of the weather questionnaire data.

Table 6. Simulation Hypotheses and Study Outcomes

Hypothesis	Supported by study results	Not supported by study results
H_1: Increased WSA from the use of portable weather applications.	√	
H_2: Using the portable weather presentation results in earlier recognition of weather and weather-state changes which will afford pilots more time to take appropriate action to avoid adverse weather.	√	
H_3: Earlier adverse weather avoidance decision-making will result in pilots maintaining their appropriate distance from the weather event.		✕
H_4: The portable device can be used without degrading pilot performance on safety-related flight tasks, actions, and decisions.	√	

44

H₁: Increased Weather Situation Awareness from the use of portable weather applications.

Using a portable weather application with selected weather information will result in much higher pilot WSA compared to piloting without the device (i.e., "see and avoid"). We assessed this hypothesis by analyzing flight trajectory and communications data to determine if there was an effect from weather application use on a participant's WSA. When looking at vertical flight profiles, we found no credible differences in altitude changes between the experimental group (portable weather application) and the control group (no weather information). However, when analyzing the horizontal flight profiles, we found a credible difference in route deviation during a convective weather scenario (Scenario A). The experimental group had credibly larger deviations from the pre-planned route compared to the control group. The experimental group had access to information that was used to plan and to make decisions whether to stay on the route or to deviate from areas of hazardous weather. These results indicate a positive effect on participants' WSA when using the portable weather presentation. An analysis of the captured transmissions related to providing weather information from aviation routine weather reports (METARs) and Terminal Area Forecasts (TAFs), as well as information related to weather-state changes acquired from the portable application, showed that the experimental group provided credibly more communications of weather information than the control group. Assessing the number of deviations to alternate airports and the scenario time at which they had occurred, we found that, compared to the control group, the experimental group made more decisions to divert and that their decisions to divert came earlier in the scenario. However, because of the low numbers, neither the number of decisions to divert nor the time at which the diversions occurred was credibly different between groups. Nevertheless, these outcomes support our hypothesis that participants had an increased WSA when using the portable weather application.

H₂: Using the portable weather presentation results in earlier recognition of weather and weather-state changes which will afford pilots more time to take appropriate action to avoid adverse weather.

Analysis of how close participants flew to areas of hazardous weather showed that the experimental group (using a portable weather application) kept larger distances from hazardous weather cells than the control group. The control group also flew credibly *closer* to 30 dBZ cells (mode = 2.59 nmi) than the experimental group (mode = 5.72 nmi). Although the experimental group kept larger distances to hazardous weather than the control group, both groups flew closer to hazardous precipitation cells (30 dBZ cells) than what is recommended in current guidelines (20 statute miles).

H₃: Earlier adverse weather avoidance decision-making will result in pilots maintaining their appropriate distance from the weather event.

The distance-to-weather analysis showed that the experimental group (using a portable weather application) kept larger distances from hazardous weather cells than the control group. However, the experimental group did not maintain an appropriate distance-to-weather as they flew too closely to 30 dBZ precipitation cells. Therefore, we failed to find empirical support for our third hypothesis.

H₄: The portable device can be used without degrading pilot performance on safety-related flight tasks, actions, and decisions.

To assess participants' cognitive engagement during scenario flights, we recorded prefrontal cortical activity using the fNIR system. Typically, the fNIR signal from neural activation is a decrease of deoxygenated hemoglobin accompanied by an increase of oxygenated hemoglobin. For Scenario A, we found a credibly higher oxygenation for the experimental group compared to the control group. For Scenario B, we found no credible differences between groups. We interpret the increased prefrontal blood oxygenation for the experimental group as symptomatic of an increased cognitive engagement due to flight planning and decision-making. This outcome is similar to what was found by Ahlstrom and Suss (2014) for pilots who detected METAR symbol changes during flight, which led to an increased level of planning and decision-making by the pilot. The outcome also supports our hypothesis that the portable weather application can be used without degrading pilot performance on safety-related flight tasks, actions, and decisions.

3.5 Flight Profiles

Our examination of flight trajectory data and communications helped us determine if there was an effect from weather application use on a pilot's WSA. In looking at vertical flight profiles, we found no credible differences in altitude changes between the experimental group (portable weather application) and the control group (no weather information). However, when analyzing the horizontal flight profiles, we found a credible difference in route deviation during a convective weather scenario (Scenario A). The experimental group had credibly larger deviations from the pre-planned route compared to the control group. The experimental group had access to information that was used to plan and to make decisions whether to stay on the route or to deviate from areas of hazardous weather. These results indicate a positive effect on participants' WSA from the use of the portable weather presentation.

3.6 Communications

An analysis of the captured transmissions related to providing weather information from METARs, TAFs, and information related to weather-state changes acquired from the portable application, showed that the experimental group provided credibly more communications of weather information than the control group. Assessing the number of deviations to alternate airports and the scenario time at which they occurred, we found that the experimental group made more decisions to divert and that their decisions to divert came earlier in the scenario. However, because of the low numbers, neither the number of decisions to divert nor the time at which the diversions occurred was credibly different between groups. Nevertheless, these outcomes support our hypothesis that participants had increased WSA when using the portable weather application.

3.7 Pilot Sensitivity to Weather Element Changes

In this study, we assessed participant sensitivity to weather symbol changes on images from the portable weather application. Using a change-detection experiment, we assessed participant discriminability of signal and noise trials using cloud ceiling, precipitation, and PIREP information. In general, participant discrimination performance was low for all conditions in comparison to the performance of a group of ideal observers.

For the discrimination of color changes in Cloud Ceiling areas, we found modest discrimination accuracy compared to the performance of a group of ideal observers, but no credible differences in discriminability between image conditions that varied in the number of color-shaded areas (100% vs. 50% area manipulations).

For the discrimination of PIREP symbols, neither the manipulation of the number of PIREP symbols (100% vs. 50%); the manipulation of the number of precipitation cells (100% and 50%) along with the PIREP symbols; nor the manipulation of the PIREP symbol intensity changes along with a manipulation of the number of precipitation cells (100% and 50%) had any effect on discrimination performance.

These findings imply (a) that participants had difficulty discriminating signal trials from noise trials and (b) that the manipulation of the 100% and 50% levels were not enough to differentiate performance. The only exception is a credible difference in response times for intensity changes of PIREP symbols between the 100% and 50% precipitation conditions, with longer response times for the 100% precipitation condition.

The outcome also implies that work is still needed to optimize the symbology for portable cockpit weather presentations. All symbol and background combinations should provide optimal luminance contrast, thereby enhancing symbol discrimination and reducing the time needed to differentiate all elements in the presentation.

3.8 Summary of Questionnaire Results

Appendix C presents an analysis of the 11 post-scenario questions. For the analysis, we used a Bayesian model (see Kruschke, 2014, p. 682) for an *ordinal predicted* variable (i.e., subjective ratings on a 1-to-7 scale) for two groups (experimental vs. control).

In general, there was no credible difference in the subjective ratings between the experimental group and the control group. However, the posterior mean group ratings for Questions 1-9 follow a trend, with higher mean ratings for the experimental group than for the control group. This is what we would predict because the experimental group had access to the portable weather application. We found a credible difference between groups for Question 8 only—which asks how easy it was to determine the potential for turbulence along the route of flight. In this case, the experimental group had access to Turbulence Potential information on their portable application, whereas the control group had no such information.

4. DISCUSSION

The outcome of this study provides empirical evidence that portable weather applications increase participant WSA. In general, this translates to 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, based on the outcome, we also see that an increased WSA does not necessarily translate to improved flight behavior. The difference in distance-to-weather during the convective scenario (Scenario A) is a good example in which the experimental group maintained greater distances away from 30 dBZ cells than the control group. Nevertheless, participants in the experimental group flew much more closely to hazardous precipitation cells than what is recommended in current FAA guidelines. What we would like to see is an improved 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.

Another example is the lack of a difference between the experimental group and the control group for Scenario B—rather than following the pre-planned route and traversing areas of reported icing and predicted reduced visibility, an improved VFR flight behavior would likely have encompassed deviations or a decision to turn around. In a sense, the outcome for the experimental group in Scenario B is similar to plan-continuation errors (Wiggins, Azar, Hawken, Loveday, & Newman, 2014), where pilots continue the flight along their pre-planned route towards their destination despite information on the portable application that suggests a deviation or a decision to turn around or land. Previous work (a) has investigated the relationship between pilots' decision to fly into IMC during VFR and the time of the onset of adverse weather and (b) has shown that pilots are more likely to fly into IMC when the weather deteriorated earlier in the flight (Wiegmann, Goh, & O'Hare, 2002).

The present results offer additional support to this finding—here, irrespective of access to the portable weather application, pilots did not deviate from their course in Scenario B, which featured weather that deteriorated early in the flight. On the other hand, pilots in the experimental group were much more likely to deviate from their course in Scenario A, which featured good VMC for the first several minutes of flight.

These outcomes suggest that pilots may also benefit from training on how to better interpret weather presentations and how to translate an increased WSA into improved 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). Of the 73 pilots that participated in the simulation, 42 pilots reported having had no additional training in weather interpretation beyond basic pilot training. Such training is crucial because mobile weather applications among GA pilots are becoming widespread.

5. CONCLUSION AND RECOMMENDATIONS

We found that a mobile weather application improves pilots' WSA and cognitive engagement. Perhaps more importantly, our results showed that with the use of a portable weather application, pilots had a greater ability to avoid areas of hazardous weather. However, the 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 FAA guidelines (i.e., 20 statute miles). Therefore, the use of a portable weather application did not translate to positive changes in the pilots' flying behavior.

From this outcome, we believe there are four recommendations that need to be addressed by future research:

1. **An assessment of the effect of pilot training on how to interpret weather information on modern electronic displays.**

   Note that in the current study, there were very few participants with prior training in how to interpret information on weather displays (two in the control group and one in the experimental group). The fact that the experimental group, on average, came as close as 5.72 nmi to yellow precipitation areas (30 dBZ intensity) directly indicates a lack of a thorough understanding of the dangers from being that close to a storm cell.

2. **An assessment of the potential effect from pilot training on how to translate weather information into enhanced flight decisions.**

   Note that in this study, participants in both groups gave the highest agreement ratings for the statement that it would make it easier to avoid hazardous weather if they used a

portable weather application during flight. However, pilots failed to incorporate their increased WSA from the use of a portable weather application into their flight decisions. This future assessment will likely also involve training on the use of electronic displays as, potentially, a large portion of current GA pilots have no or very little experience with electronic displays.

3. **We also recommend research for future weather applications that explores other ways to provide clear display indications of areas to avoid during flight.**

   Here, there are many alternatives to display, such as hazardous precipitation areas. Instead of indicating areas of varying NEXRAD intensities, precipitation displays could indicate all areas within 20 statute miles that should be avoided.

4. **Finally, the change-detection experiment shows that participants have difficulty to discriminate some symbol and background combinations.**

   This should be addressed by future research to provide portable weather applications with optimal luminance contrast between weather elements and the backgrounds, thereby enhancing symbol discrimination and reducing the time needed to differentiate all elements in the presentation.

References

Ahlstrom, U., & Dworsky, M. (2012). *Effects of weather presentation symbology on general aviation pilot behavior, workload, and visual scanning* (DOT/FAA/TC-12/55). Atlantic City International Airport, NJ: FAA William Hughes Technical Center.

Ahlstrom, U., & Suss, J. (2014). *Now you see me, now you don't: Change blindness in pilot perception of weather symbology* (DOT/FAA/TC-14/16). Atlantic City International Airport, NJ: FAA William Hughes Technical Center.

Federal Aviation Administration. (2010). *Weather technology in the cockpit program capabilities report* (DTFAWA-09-C-00088). Norman, OK: Atmospheric Technology Services Company, LLC.

Federal Aviation Administration. (2013). *Thunderstorms* (Advisory Circular 00-24C). Washington, DC: Federal Aviation Administration.

Federal Aviation Administration, & National Oceanic and Atmospheric Administration. (1983). *Thunderstorms* (DOT/FAA/AC 00–24B). Oklahoma City, OK: FAA and NOAA.

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* (2nd ed.). Academic Press/Elsevier. ISBN: 9780124058880.

Lee, M. D. (2008). BayesSDT: Software for Bayesian inference with signal detection theory. *Behavior Research Methods*, *40*(2), 450–456.

Olman, C., & Kersten, D. (2004). Classification objects, ideal observers, and generative models. *Cognitive Science, 28*, 227–239.

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: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

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

Ware, C., & Arsenault, R. (2012). Target finding with a spatially aware handheld chart display. *Human Factors 54*(6), 1040–1052.

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decision to continue visual flight rules flight into adverse weather. *Human Factors, 44*(2), 189–197.

Wiggins, M., & O'Hare, D. (2003). Weatherwise: Evaluation of a cue-based training approach for the recognition of deteriorating weather conditions during flight. *Human Factors, 45*, 337–345.

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.

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

Acronyms

FAA	Federal Aviation Administration
fNIR	Functional Near-Infrared
GA	General Aviation
HDI	High Density Interval
IFR	Instrument Flight Rules
IMC	Instrument Meteorological Conditions
MCMC	Markov Chain Monte Carlo
METAR	Meteorological Aerodrome Report
MVMC	Marginal Visual Meteorological Conditions
PIREP	Pilot Report
RGB	Red, Green, Blue Color Space
ROPE	Region of Practical Equivalence
SD	Signal Detection
SDT	Signal Detection Theory
SES	Stimulus Experiment System
SME	Subject Matter Expert
TAF	Terminal Area Forecast
VFR	Visual Flight Rules
VMC	Visual Meteorological Conditions
VOR	Omnidirectional Radio Range
WMA	Windows Media Audio
WSA	Weather Situation Awareness
WTIC	Weather Technology in the Cockpit

Appendix A: Informed Consent Form

Informed Consent Statement

I, _____, understand that this pilot study, entitled "Weather Technology in the Cockpit" is sponsored by the Federal Aviation Administration (FAA) and is being directed by Ulf Ahlstrom.

Nature and Purpose:

I volunteered as a participant in this study that encompasses two cockpit simulation flights and two symbol detection experiments. The primary purpose of the symbol detection experiments and the cockpit simulations are to improve General Aviation (GA) weather presentations for the cockpit. During the experiments, participants will evaluate sets of static images displayed on a computer monitor. During the simulation, participants will fly a single-engine GA simulator during Visual Flight Rules (VFR) conditions while avoiding encounters with hazardous weather.

Research Procedures:

Eighty GA pilots will participate as volunteers during a half-day (4 hours) that covers two simulation flights and two experiments. The participants will be engaged from 8:00 AM to 12:00 PM (or from 12:00 PM to 16:00 PM) with short rest breaks. All the participants will conduct the simulator flights before performing the experiments.

The first part of the session will encompass a briefing to review project objectives and participant rights and responsibilities. This briefing will also include initial familiarization training on the cockpit simulator, weather presentations, and the fitting of a head-mounted, Functional Near-Infrared Spectroscopy (fNIR) sensor used to measure cognitive workload. The participant will first complete a practice flight scenario. After the practice scenario, the participant will fly a designated route during a simulator flight (approximate duration: 40 minutes). During the simulator flight, an automated data-collection system will record cockpit system operations and generate a set of standard cockpit simulation measures including communications.

After the simulation flight, the participants will complete a questionnaire to report their overall workload, situation awareness, and provide an assessment of the cockpit system and test conditions. Additionally, participants will complete a brief biographical background questionnaire covering their flight and weather display experience.

After completing the questionnaires, participant will be briefed on the image detection experiment and thereafter conduct a training session. After the training session participants will complete the experimental task, which is divided into blocks. During these blocks, an automated data collection system records each participant response.

Anonymity and Confidentiality:

The information that I provide as a participant is strictly confidential and I shall remain anonymous. I understand that no Personally Identifiable Information [PII] will be disclosed or released, except as may be required by statute. I understand that situations when PII may be disclosed are discussed in detail in FAA Order 1280.1B "Protecting Personally Identifiable Information [PII]."

Benefits:

I understand that the only benefit to me is that I will be able to provide the researchers with valuable feedback and insight into weather presentation symbology. My data will help the FAA to establish human factors guidelines for weather displays and assess if there is a need to standardize the symbology for enhanced weather information.

Participant Responsibilities:

I am aware that to participate in this study I must be a GA pilot. I am also aware that I am not allowed to participate if I have a personal and/or familial history of epilepsy.

I will (a) fly the designated route in the cockpit simulations, (b) perform the experiments, and (c) answer questions asked during the study to the best of my abilities. I will not discuss the content of the part-task or the cockpit simulation with other potential participants until the study is completed.

Participant Assurances:

I understand that my participation in this study is voluntary, and I can withdraw at any time without penalty. I also understand that the researchers in this study may terminate my participation if they believe this to be in my best interest. I understand that if new findings develop during the course of this research that may relate to my decision to continue participation, I will be informed. I have not given up any of my legal rights or released any individual or institution from liability for negligence.

The research team has adequately answered all the questions I have asked about this study, my participation, and the procedures involved. I understand that Ulf Ahlstrom or another member of the research team will be available to answer any questions concerning procedures throughout this study. If I have questions about this study or need to report any adverse effects from the research procedures, I will contact Ulf Ahlstrom at (609) 485-8642.

Discomfort and Risks:

The fNIR sensor, consisting of a silicon pad containing small light-emitting diodes (LEDs) and light detectors, will be placed over the participant's forehead with a headband. Low power light will be shone onto the forehead area during the simulator flight, and changes in the amount of light that returns to the light detectors will be used to calculate changes in the concentrations of oxygenated and deoxygenated hemoglobin in the blood. The risk associated with using the fNIR sensor is less than the risk associated with spending an equivalent amount of time in the United States sunlight without wearing a hat. If the fNIR sensor causes discomfort, please alert the experimenter immediately.

In the part-task experiment, the screen may flicker back-and-forth between two images, at the rate of several times per second. For healthy individuals, there are no reported adverse effects of this common presentation technique. However, such flickering could cause seizures in epileptics. If you experience any discomfort due to the presentation of the images, please alert the experimenter immediately.

I agree to immediately report any injury or suspected adverse effect to Ulf Ahlstrom at (609) 485-8642.

Signature Lines:

I have read this informed consent form. I understand its contents, and I freely consent to participate in this study under the conditions described. I understand that I may request a copy of this form.

Research Participant:_____   Date:_____

Investigator:_____   Date:_____

Witness:_____   Date:_____

Appendix B: Biographical Questionnaire

Biographical Questionnaire

Demographic Information and Experience

1. What pilot certificate and ratings do you hold? (circle as many as apply)	Private Commercial ATP Glider SEL SEA MEL Airship Instrument CFI CFII MEI Helicopter A&P IA

2. What is your age?   _____ Years

3. Approximately, what is your total time?   _____ Hours

4. Approximately how many actual instrument hours do you have?   _____ Hours

5. Approximately how many instrument hours have you logged in the last 6 months   _____ Hours
    (simulated and actual)?

1. List all (if any) in-flight weather presentation systems you have used during a flight to make actual weather judgments (not including onboard radar or Stormscope).

    _____

    _____

2. Have you had any training in weather interpretation other than basic pilot training (for example, courses in meteorology)? If so, to what extent?

    _____

    _____

3. How often *do you provide/did you provide* pilot reports (PIREPs) during actual GA flights?

    _____

    _____

Thank you very much for participating in our study, we appreciate your help.

Appendix C: Post-Scenario Questionnaire with Analysis

Post-Scenario Questionnaire with Analysis

1. To what degree did weather conditions affect your decision to deviate from your pre-planned course?

None At All						Very Much
1	2	3	4	5	6	7
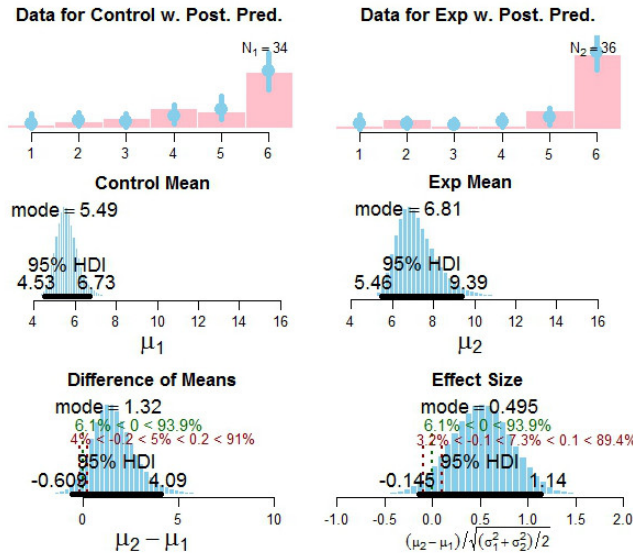
2. How easy was it to avoid areas of Instrument Meteorological Conditions (IMC)?

Very Hard						Very Easy
1	2	3	4	5	6	7
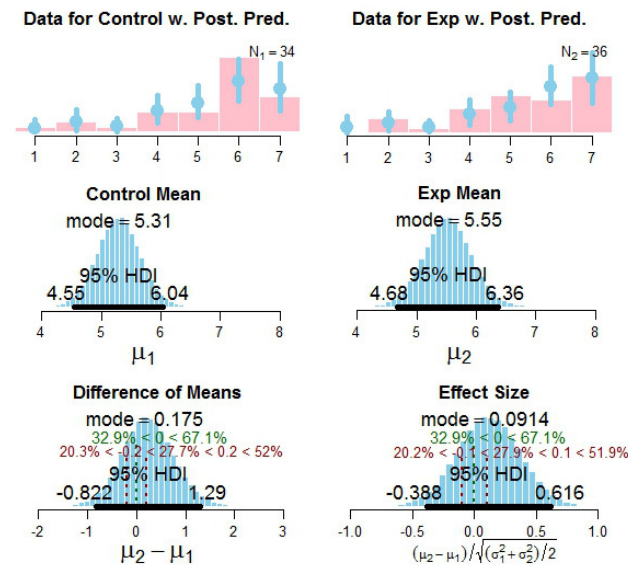
3. How easy was it to determine the location of severe precipitation areas?

Very Difficult						**Very Easy**
1	**2**	**3**	**4**	**5**	**6**	**7**
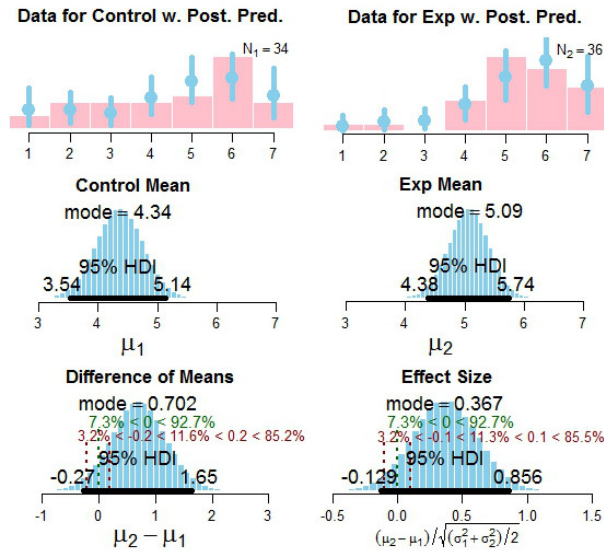
4. How easy was it to determine the distance from the aircraft to areas of precipitation?

Very Difficult						**Very Easy**
1	**2**	**3**	**4**	**5**	**6**	**7**
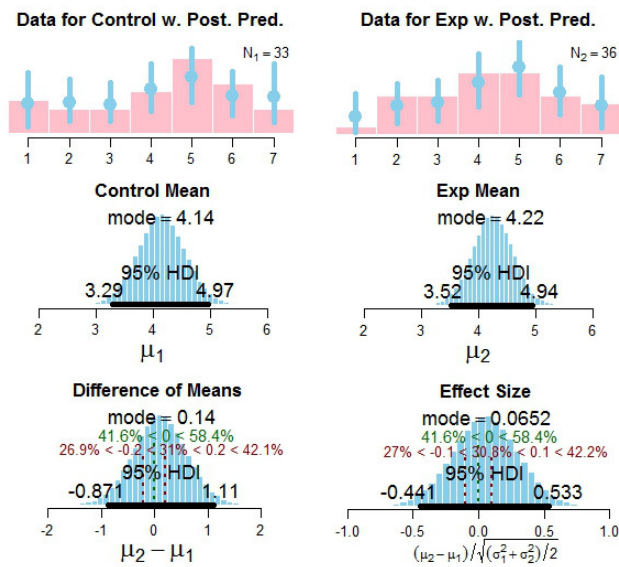
5. How easy was it to determine areas of poor visibility?

Very Hard						**Very Easy**
1	**2**	**3**	**4**	**5**	**6**	**7**

Data for Control w. Post. Pred. ($N_1 = 34$)
Data for Exp w. Post. Pred. ($N_2 = 36$)

Control Mean mode = 4.77, 95% HDI 4.09 5.46, μ_1
Exp Mean mode = 5.16, 95% HDI 4.43 5.85, μ_2

Difference of Means mode = 0.365
20.7% < 0 < 79.3%
10.1% < -0.2 < 26% < 0.2 < 64%
95% HDI -0.509 1.24
$\mu_2 - \mu_1$

Effect Size mode = 0.214
20.7% < 0 < 79.3%
11.1% < -0.1 < 22.6% < 0.1 < 66.3%
95% HDI -0.285 0.692
$(\mu_2 - \mu_1)/\sqrt{(\sigma_1^2 + \sigma_2^2)/2}$
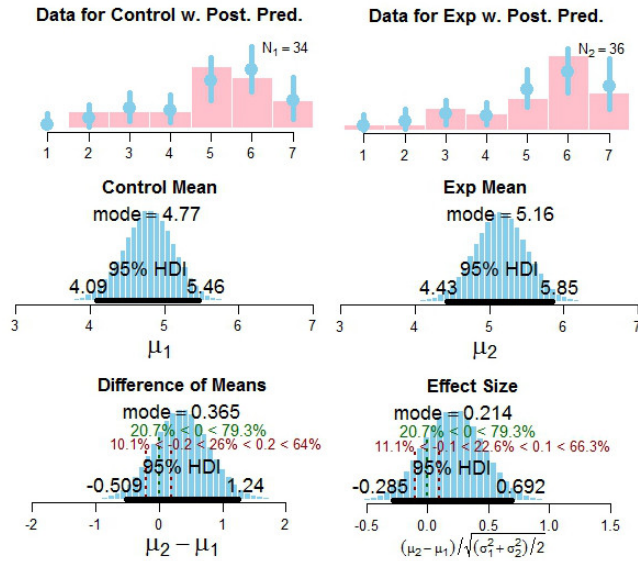
6. How easy was it to determine cloud ceilings?

Very Hard						**Very Easy**
1	**2**	**3**	**4**	**5**	**6**	**7**

Data for Control w. Post. Pred. ($N_1 = 34$)
Data for Exp w. Post. Pred. ($N_2 = 36$)

Control Mean mode = 4.01, 95% HDI 3.21 4.85, μ_1
Exp Mean mode = 4.9, 95% HDI 4.22 5.57, μ_2

Difference of Means mode = 0.87
3.5% < 0 < 96.5%
1.3% < -0.2 < 6.6% < 0.2 < 92.1%
95% HDI -0.073 1.84
$\mu_2 - \mu_1$

Effect Size mode = 0.47
3.5% < 0 < 96.5%
1.3% < -0.1 < 6.5% < 0.1 < 92.2%
95% HDI -0.0369 0.943
$(\mu_2 - \mu_1)/\sqrt{(\sigma_1^2 + \sigma_2^2)/2}$
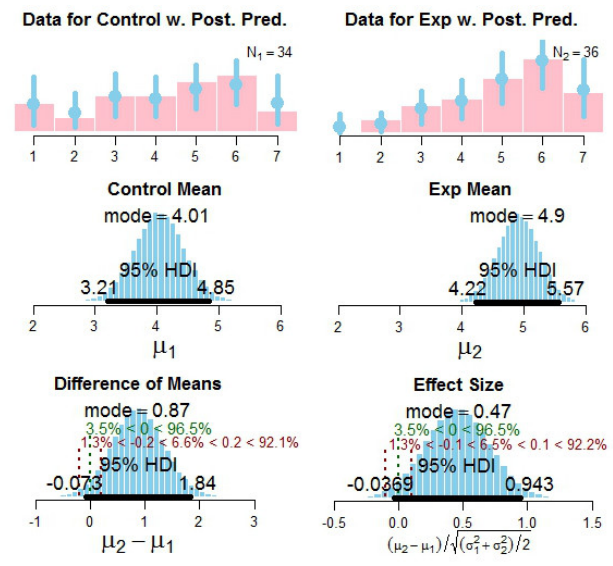
7. How easy was it to determine the potential for icing conditions along the route of flight?

Very Hard						**Very Easy**
1	**2**	**3**	**4**	**5**	**6**	**7**

Data for Control w. Post. Pred.
$N_1 = 34$

Data for Exp w. Post. Pred.
$N_2 = 36$

Control Mean
mode = 4.18
95% HDI
3.23 5.13
μ_1

Exp Mean
mode = 4.77
95% HDI
4.06 5.41
μ_2

Difference of Means
mode = 0.541
14.8% < 0 < 85.2%
8% < -0.2 < 16.8% < 0.2 < 75.2%
95% HDI
-0.533 1.63
$\mu_2 - \mu_1$

Effect Size
mode = 0.267
14.8% < 0 < 85.2%
7.5% < -0.1 < 18.3% < 0.1 < 74.2%
95% HDI
-0.234 0.756
$(\mu_2 - \mu_1)/\sqrt{(\sigma_1^2 + \sigma_2^2)/2}$

8. How easy was it to determine the potential for turbulence along the route of flight?

Very Hard						**Very Easy**
1	**2**	**3**	**4**	**5**	**6**	**7**

Data for Control w. Post. Pred.
$N_1 = 34$

Data for Exp w. Post. Pred.
$N_2 = 36$

Control Mean
mode = 2.99
95% HDI
2.11 3.89
μ_1

Exp Mean
mode = 4.48
95% HDI
3.74 5.21
μ_2

Difference of Means
mode = 1.46
0.3% < 0 < 99.7%
0.1% < -0.2 < 0.8% < 0.2 < 99.1%
95% HDI
0.421 2.56
$\mu_2 - \mu_1$

Effect Size
mode = 0.693
0.3% < 0 < 99.7%
0.1% < -0.1 < 0.9% < 0.1 < 99%
95% HDI
0.191 1.19
$(\mu_2 - \mu_1)/\sqrt{(\sigma_1^2 + \sigma_2^2)/2}$

C-4

9. How would you rate your mental workload during the flight?

Very Low						Very High
1	**2**	**3**	**4**	**5**	**6**	**7**
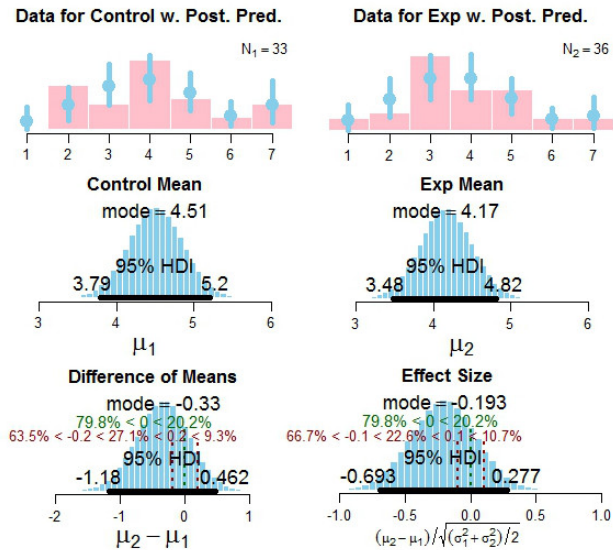
10. Having a hand-held weather presentation during flight would make it easier to avoid hazardous weather compared to only looking out the cockpit window.

Totally Disagree						Totally Agree
1	**2**	**3**	**4**	**5**	**6**	**7**
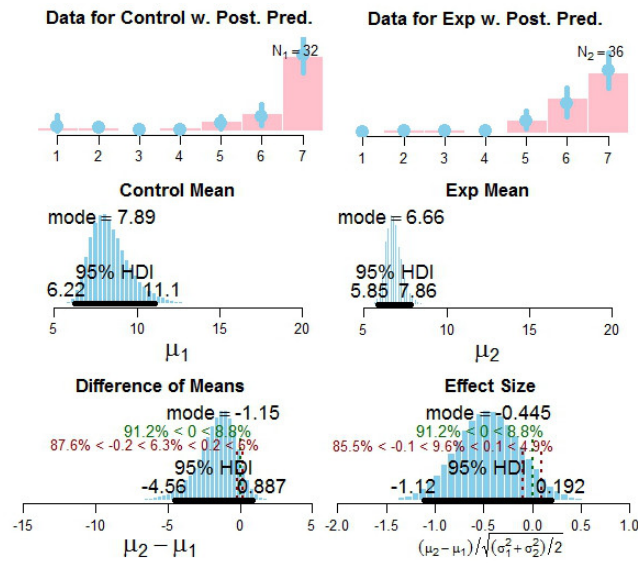
11. Did you experience discomfort from the fNIR device?   No   Yes

No = 67
Yes = 3

Appendix D: Research Staff List

Research Staff List

Name	Role	Responsibility
Ahlstrom, Ulf	Test Lead	Manages the project: Lead developer of scenarios, change-detection experiments, test plan, data analysis, and Technical Report. Test conductor.
Bastholm, Robert	Human Factors Specialist	Implements fNIR recordings. Prepares training materials, data analysis, and Technical Report. Test conductor.
Caddigan, Eamon	Human Factors Specialist	Co-develops test plan, data analysis, Technical Report. Test conductor lead.
Dworsky, Matthew	Human Factors Specialist	Implements fNIR recordings. Develops change-detection stimuli, prepares training materials, test conductor. Data analysis, Technical Report.
Granich, Thomas	Software Engineer	Implements simulator system and cockpit data recordings.
Jackman, April	Technical Editor	Reviews, edits, formats, and prepares Technical Reports for publication and dissemination.
Johnson, Ian	WTIC Human Factors Lead	Provides pilot perspective. Coordinates on test plan and test efforts, technical review of test products and deliverables.
Kukorlo, Matt	Pilot Subject Matter Expert	Provides pilot perspective. Flight Scenario Developer. Simulation SME.
Kusza, Robert	Simulator Developer	Operates simulator systems, records cockpit data, and performs data backup.
Mutchler, Mark	Certification and small aircraft expertise	Provides pilot perspective. Assist in scenario development.
Ohneiser, Oliver	Human Factors Specialist, Software Engineer	Provides analysis software development. Technical Report.
Pokodner, Gary	WTIC Program Manager	Tracks project, conducts interim reviews, and final acceptance of deliverables.
Rehman, Al	Manager, Cockpit Simulator Lab	Manages and maintains cockpit lab systems. Provides support in scheduling subject pilots for study.
Sultan, Roger	Safety	Provides pilot perspective. Assists with scenario development. Identifies any safety or documentation issues. Provides aeronautical standards and expertise.