JOURNAL OF AIR TRANSPORTATION
Vol. 29, No. 4, October–December 2021

General Aviation Pilots' Capability to Interpret Aviation Weather Displays

Beth Blickensderfer,* Jacqueline McSorley,[†] Nicolas Defillipis,[†] Jayde M. King,[†] Yolanda Ortiz,[†]
Thomas A. Guinn,[‡] and Robert Thomas[§]
Embry-Riddle Aeronautical University, Daytona Beach, Florida 32114

https://doi.org/10.2514/1.D0242

Current literature indicates that a lack of weather knowledge and poor product interpretability may be contributing to the high probability of fatalities in general aviation weather-related accidents. Eight hundred and thirty-seven general aviation pilots completed an online aviation weather product interpretation test that asked pilots to apply information gleaned from weather hazard products for fight planning. Participants were divided into five categories of certificate/ratings. A total of 118 questions were divided into five separate tests and randomly distributed to the participants. A series of analyses were conducted to assess the impact of weather product and pilot certification on interpretation scores. Overall, certified private pilots scored significantly lower than certificated commercial pilots, flight instructors, and airline transport pilots. Private with instrument rating pilots scored significantly lower than certificated flight instructors and air transport pilots. Further analysis revealed that pilots scored lowest on ceiling visibility analysis, satellite, station plots, and surface prognostic products. Highest scores were associated with winds aloft, graphical turbulence, and pilot reports. The results have implications for both weather display design and pilot training.

I. Introduction

A PREVAILING problem in general aviation (GA) is the fatality rate in weather-related accidents [1]. One concern is flights operating under visual flight rules (VFR) that inadvertently enter instrument meteorological conditions (IMC) [1]. Despite the development of new and updated weather displays, GA pilots can unexpectedly encounter hazardous weather that affects the safety of their flights. Within the last decade, the National Transportation Safety Board named communication of weather to GA pilots one of the most wanted improvements in transportation in multiple reports [1–5]. In simple terms, "identifying" weather phenomena (including hazardous weather) is accomplished by meteorology systems and personnel [6]. Once weather phenomena have been identified, the next step is "communicating" this information to a user (in this case, GA pilots). This requires a system (or person) to send a message (i.e., the weather information display), with the goal being that users (i.e., pilots) will receive and interpret the information as intended [7]. Thus, one step to assessing communication of weather to GA pilots, and the goal of the current paper, is to determine the degree to which GA pilots can interpret existing weather information displays (i.e., weather communications).

A. Aviation Weather Products

GA incurs more fatalities per weather-related accident than commercial or military aviation [8], and these accidents come with a high likelihood of fatality [1]. Both comprehensive reviews of accident data and individual research studies demonstrate that a problem exists. For example, according to the most recent Nall Report, approximately 20% of GA fatalities stem from weather-related accidents [1]. Further, in a recent study of real-time flights, Boyd captured GA aircraft circumventing extreme convection with 69% (en route) and 93% (landing) violating the Federal Aviation Administration (FAA)–recommended 20 nautical miles separation distance from extreme convection [9]. Ultimately, hazardous weather significantly increases the chance of a fatal accident when present, particularly in small, low-ceiling aircraft [10].

To mitigate these accidents, it is important to examine the factors that could lead up to such an accident, including the interpretation of aviation weather products. Aviation weather information displays, referred to as aviation weather products, include weather forecasts, weather observations, and advisory information presented in various manners. Forecast products display a projected outlook of future weather (e.g., terminal aerodrome forecasts [TAFs], winds aloft, and others). Observation products depict current weather conditions at or near airports and beyond (e.g., meteorological aerodrome reports [METARs], satellite, and radar). Advisory products provide displays of current and/or forecast severe and moderate weather conditions that pilots should avoid (e.g., significant meteorological information [SIGMETs] and graphical airmen's meteorological information [G-AIRMET]). One purpose of the displays is to assist GA pilots to map out potential weather phenomena along their planned flight route. The FAA recommends a 3-P model to guide GA pilots' preflight weather planning and in-flight decision making: *perceive, process*, and *perform* [11]. *Perceive* includes looking for hazards that could harm the flight in a variety of weather products. *Process* consists of interpreting the information received by the weather products to form a big picture of the potential weather. *Perform* requires pilots to act in a way that avoids or diminishes potentially dangerous elements. When examining the weather products, if pilots are not able to develop an accurate understanding of the weather (i.e., perceive and process) and/or plan weather-appropriate flight routes (i.e., perform), the door may be open for those pilots to encounter unexpected hazards.

B. Evidence of Weather Display Interpretation Challenges

Current literature indicates that a gap exists between what the weather products aim to broadcast and what pilots glean from it [12–15]. One area of research interest has been the effectiveness of Next Generation Weather Radar (NEXRAD) to communicate aviation weather hazards to GA pilots. NEXRAD "makes conventional reflectivity observations and also uses the 'Doppler effect' to

---

Received 14 October 2020; revision received 21 April 2021; accepted for publication 28 May 2021; published online 22 July 2021.

*Professor of Human Factors, Department of Human Factors and Behavioral Neurobiology.

[†] Graduate Research Assistant, Department of Human Factors and Behavioral Neurobiology.

[‡] Professor of Meteorology, Department of Applied Aviation Science.

[§] Assistant Professor of Aeronautical Science, Department of Aeronautical Science.

measure motion of clear air and atmospheric phenomena within storms" [16]. Beginning nearly two decades ago, research has indicated that pilots misinterpret convective weather scenarios. First, Latorella and Chamberlain [13] assessed graphical weather information systems (GWISs) in order to determine potential usability issues in these systems. This research discovered that pilots were using GWISs to fly closer to hazardous weather. In a follow-up study, Latorella and Chamberlain [17] found that higher-resolution NEXRAD displays prompted pilots to fly closer to areas of bad weather in order to "tactically" avoid it. Additionally, Beringer and Ball [14] assessed the effects of NEXRAD display resolution on pilots' decision to fly into weather. Similar to Latorella and Chamberlain [13,17], they found that higher-resolution displays prompted pilots to fly closer to severe weather due to higher confidence in being able to narrowly avoid the weather. The pilots did not factor in system latency and short-notice weather changes in their judgments of weather severity. A decade later, Knecht [18] also reported deficiencies in the capability of NEXRAD displays to communicate information to GA pilots. This work emphasized the importance of implementing future weather predictions and a range ring in order to improve judgments of closest point approach. From a training perspective, research has demonstrated gaps in pilots' understanding of NEXRAD, but pilots did benefit from short training courses [19,20].

Many other weather products and displays exist beyond NEXRAD, and these also warrant investigation. In a detailed analysis of weather-related accident case studies, Lanicci et al. [10] determined that, from a meteorological standpoint, the information provided by the weather products at the time of the accidents was accurate. The GA pilots, however, did not seem to fully understand how to access or apply the available information. Gaps in pilot understanding included the following weather hazards: IMC, convective weather, icing, turbulence, and wind shear. An implication of the Lanicci et al. study is that weather products may be difficult for pilots to interpret correctly (i.e., a communication gap may exist between the products and the GA pilot user group). Similarly, Wiggins [21] assessed pilot assessments of weather phenomena in deteriorating conditions. Wiggins found that situation awareness and pilot experience are important factors that affect weather interpretation. This study also identified long-term memory and situational cues as essential ways to bridge the information gap; however, if pilots do not have the proper training and weather knowledge in the first place, this could affect the utilization of these cues [21].

Evidence of a lack of interpretability of weather displays by GA pilots is mounting. Burian and Jordan [12] conducted a study to determine how well pilots understood weather products in regard to flight-related decision making. Weather surveys were given to pilots and certificated flight instructors (CFIs) to assess general weather knowledge. Self-reports indicated that the majority of participants believed that they had a moderate understanding of aviation weather displays, but actual test results showed a distinct lack of understanding. This finding points toward the idea that pilots have difficulty interpreting aviation weather products. Recently, Blickensderfer and colleagues conducted a broad assessment of weather product interpretability [22,23]. The researchers developed a 95-question multiple-choice test (the Aviation Weather Product Test) to quantify how well low-hour GA pilots could answer aviation weather interpretability problems. Scores, on average, were moderately low. Student pilots scored significantly lower than all other groups. Commercial certificated pilots with instrument ratings performed the best, but still only averaged 65% correct answers across the whole test. Regardless of certificate/rating, GA pilots scored higher on upper level charts, convective SIGMETs, and surface analysis charts in comparison to the rest of the hazard products. GA pilots' scores were lowest on textual METARS, as well as radar and satellite questions. As this study used a sample of college students, it is unknown how experienced GA pilots can interpret weather products. To increase the validity of the findings, a sample of participants that is more generalizable to the GA pilot population is needed.

C. Current Study

Aviation literature illuminates a high fatality rate in GA weather-related accidents [1], as well as performance decrements despite meteorologically accurate weather products [10]. Additional research demonstrated that interpretation problems exist across a broad suite of weather products, particularly among the lower-rated, lower-flight-hour GA population [22]. The purpose of the current study was to examine GA pilots' capability to interpret aviation weather displays/products in a more generalizable sample that reflects the typical GA pilot in terms of age and number of flight hours.

II. Method

A. Participants

Participants ($n = 837$) were certificate-holding pilots aged 18–86 years (mean $[M] = 57.0$ years, standard deviation $[SD] = 13.7$ years). The pilots were recruited through the use of the Aircraft Owners and Pilots Association's (AOPA's) member e-mail listserv. Participation was voluntary and, as an incentive, participants were offered to be entered into a drawing for a small prize package. Although 1702 participants began the survey, only those who completed the entire survey were included in data analysis. All pilots held certificates in the following: private pilot ($n = 248$), private with instrument rating ($n = 240$), commercial pilot with instrument rating ($n = 134$), CFI ($n = 115$), or airline transport pilot (ATP) ($n = 100$). Table 1 displays the descriptive statistics for flight hours by pilot certificate/rating. The Embry-Riddle Aeronautical University Institutional Review Board approved this study for the protection of human participants.

B. Measures

The two key measures of this study, the demographics questionnaire and Aviation Weather Product Test, were implemented via the online survey system Qualtrics [23].

1. *Demographic Questionnaires*

The demographic questionnaire consisted of 15 items. The items obtained basic information about the participants such as age, flight experience, and weather training (e.g., frequency of weather product use and details on where they received this training).

2. *Aviation Weather Product Test*

The purpose of the 118-question Aviation Weather Product Test was to evaluate GA pilots' capability to interpret weather products. This test exhibits high cognitive fidelity with an emphasis on the application questions. All questions were multiple-choice, and each had 3–4 answer options (i.e., a, b, and c, or a, b, c, and d), with one correct answer per question. This included 95 questions from the Blickensderfer et al.'s [22] weather interpretation assessment as well as 23 additional questions. The same research team that developed the original questions (consisting of one meteorologist, one Gold Seal–certificated flight instructor, an industrial-organizational psychologist, and two human factors specialists) developed the 23 new items. Information from the 23 new items stemmed from the FAA Advisory Circular 00-45H Change 1 [24] and Title 14 of the Code of Federal Regulations [25].

To reduce the number of questions any one pilot would be asked to answer, the 118 questions were divided into five separate tests with 20–25 questions in each. The tests were organized by topic and/or

Table 1 Descriptive statistics for flight hours by pilot certificate/rating

Weather product/ topic	Private pilot	Private with instrument	Commercial with instrument	CFI	ATP	Total
N	248	240	134	115	100	837
M	505.56	1389.08	2367.88	3568.18	8769.50	2611.21
SD	646.4	1147.4	2345.2	2943.2	6067.6	3847.60
Mdn	290	1000	1500	2550	7000	1100

weather product such that all questions pertaining to a specific weather product were presented together on a test. Test 1 contained data source, significant weather, storm definition, and product attributes questions. Test 2 contained METAR, pilot report (PIREP), winds aloft, and TAF questions. Test 3 contained current icing potential (CIP), graphical airman's meteorological information (G-AIRMET), and graphical turbulence guidance (GTG) questions. Test 4 contained radar, SIGMET, and thunderstorm concept questions. Lastly, Test 5 contained questions on the ceiling visibility analysis (CVA), satellite, station plots, and surface prognostic products.

Below are two examples of the questions presented to participants, where the bolded responses represent correct answers:

Prior to departure, it is helpful to conduct a preflight self-briefing. The goal of the self-briefing should be to:

1) Eliminate the need for an FSS standard briefing.
2) **Develop an overall mental picture of current and forecast weather conditions**.
3) Determine the safest navigation route through areas of potential widespread thunderstorms.
4) Identify the current airfield weather conditions for takeoff and landing.

You are flying to Oklahoma City, Oklahoma (KOKC) in July. You are concerned about thunderstorms. Using the below METAR, which of the following statements is true?

KOKC 150953Z 29010KT 5SM -RA SCT017 OVC030 21/19 A2998 RMK AO2 P0001 TS DSNT W MOV E

1) **Thunderstorms are moving towards the airport**.
2) Thunderstorms are moving away from the airport.
3) Thunderstorms are reducing the visibility to 5 SM.
4) There are no thunderstorms currently being reported.

For each pilot, the percentage correct was calculated for the whole test as well as for each individual category within that test. For example, if a pilot took Test 1, he/she received an overall percentage score for Test 1 as well as percentage scores for Data Sources, Significant Weather, Storm Definitions, and Product Attributes, respectively.

C. Procedure

When participants agreed to participate in the study, they were randomly assigned to receive a particular test. They accessed the test through the survey link in the original e-mail. Participants completed the test on their personal electronic devices in a location of their choosing. The devices may have included laptops, desktops, phones, and tablets. Upon receiving the e-mail with the survey link, the participant clicked on the link to open the survey. Participants read the consent form, and if in agreement to continue, they clicked "I Agree," which began the survey. Following the consent form, participants filled out the demographic questionnaire. Demographic questions were first, and the Aviation Weather Product Test questions were second. Participants were not restricted on time, and they could exit/pause the survey and continue later, as long as they used the same device. At the end of the survey, participants were invited to provide their e-mail address to be entered into a drawing to win the prize package. There was one prize package drawing for each of the five tests.

III. Results

The IBM Statistical Package for the Social Sciences (SPSS) version 25 was used to determine equivalency of groups, aggregated results, and group differences in ability to interpret weather information [26]. Descriptive statistics are shown in Tables 2–7.

A. Equivalency of Groups

To determine equivalency of the groups (e.g., participants who took Test 1 versus Test 2 and so on), mean flight hours were first examined. A 5×5wo-way, between-groups analysis of variance was conducted to explore the relationship between test number (Test 1, Test 2, Test 3, Test 4, and Test 5) and pilot rating (private, private with instrument rating, commercial with instrument rating, CFI, and ATP) on flight hours. There was a significant main effect for pilot certificate/rating on flight hours [$F(4{,}850) = 196.99$, $p < 0.001$, partial $\eta^2 = 0.48$]. A Bonferroni post hoc comparison indicated that, regardless of the test taken, private pilots ($M = 505.56$, $SD = 646.4$, median [Mdn] = 290.0) had significantly fewer flight hours than all other ratings ($p < 0.001$). Private with instrument-rated pilots ($M = 1389.1$, $SD = 1147.4$, $Mdn = 1000.0$) had significantly fewer flight hours than commercial with instrument ($p = 0.04$), ATP ($p < 0.001$), and CFIs ($p < 0.001$). Commercial with instrument-rated pilots ($M = 2367.9$, $SD = 2345.2$, $Mdn = 1500.0$) had significantly fewer than CFI ($M = 3568.2$, $SD = 2943.2$, $Mdn = 2550.0$; $p = 0.005$) and ATP-rated pilots ($M = 8769.5$, $SD = 6067.6$, $Mdn = 7000.0$; $p < 0.001$). There was not a significant main effect of test number on flight hours [$F(4{,}850) = 0.51$, $p = 0.73$, partial $\eta^2 = 0.002$], and there was not a significant interaction between test number and pilot rating on flight hours [$F(16{,}850) = 1.07$, $p = 0.38$, partial $\eta^2 = 0.02$]. This indicates an equivalency of groups who took the five tests.

B. Overall Scores Across Tests

The descriptive statistics for score by test and pilot certificate/rating are shown in Table 2.

A two-way between-groups analysis of variance was conducted to explore the impact of pilot certificate/rating (private, private with instrument rating, commercial with instrument rating, CFI, and ATP) and test number (Test 1, Test 2, Test 3, Test 4, and Test 5) on test score. A significant main effect for pilot certificate/rating occurred [$F(4{,}857) = 12.48$, $p < 0.001$, partial $\eta^2 = 0.55$]. Bonferroni post hoc comparisons indicated that, regardless of test taken, private pilots scored significantly lower than commercial with instrument, ATP, and CFI pilots ($p \leq 0.001$). Scores for private with instrument-rated pilots were significantly lower than CFI ($p = 0.02$) and ATP pilots ($p = 0.005$). Scores for commercial with instrument pilots were not significantly lower than ATP or CFI. Scores for ATP ($M = 72.6$, $SD = 14.1$) and CFI-rated pilots were not significantly different. A significant main effect of test also occurred [$F(4{,}857) = 53.39$, $p < 0.001$, partial $\eta^2 = 0.20$]. Post hoc comparisons indicated that, regardless of certificate/rating, scores were significantly higher on Test 1 than all other tests ($p < 0.001$). Scores were significantly lower on Test 5 than all other tests ($p < 0.001$). No other significant differences were found between the tests. The interaction between pilot certificate/rating and test was not significant

Table 2 Descriptive statistics for score (percentage correct) by test and pilot certificate/rating

Test	Private pilot ($n = 248$), $M(SD)$	Private with instrument ($n = 240$), $M(SD)$	Commercial with instrument ($n = 134$), $M(SD)$	CFI ($n = 115$), $M(SD)$	ATP ($n = 100$), $M(SD)$	Total ($n = 837$), $M(SD)$
1	76.1 (12.6)	81.3 (10.1)	81.2 (11.9)	81.4 (13.2)	83.4 (11.6)	79.5 (11.5)
2	61.2 (12.8)	65.4 (12.7)	65.1 (16.2)	71.7 (12.3)	75.5 (14.3)	66.0 (14.6)
3	66.2 (14.4)	65.7 (14.7)	75.0 (15.4)	73.2 (12.3)	71.2 (11.7)	67.7 (14.8)
4	60.1 (9.9)	64.5 (12.9)	69.8 (11.3)	70.2 (12.1)	69.0 (12.1)	64.9 (12.3)
5	55.4 (14.8)	61.9 (16.5)	58.8 (19.0)	59.8 (13.3)	65.2 (14.8)	59.4 (16.4)
Total	64.7 (14.3)	67.3 (15.1)	70.0 (16.9)	72.7 (14.3)	72.57 (14.1)	67.9 (15.5)

[$F(16{,}774) = 1.35$, $p = 0.157$, partial $\eta^2 = 0.027$]. This indicates that the performance trend across the different tests was approximately the same for each pilot certificate/rating group.

C. Test 1 Topics Analysis

Test 1 consisted of four content/topic areas (Data Sources, Significant Weather, Storm Definitions, and Product Attributes) with 5 questions each for a total of 20 questions. The descriptive statistics for Test 1 are shown in Table 3.

A mixed (between and within) analysis of variance assessed the impact of pilot certificate/rating and topic within Test 1 on test score. A significant main effect occurred for pilot certificate/rating on Test 1 score [$F(4{,}191) = 2.96$, $p = 0.02$, partial $\eta^2 = 0.06$]. Using Bonferroni post hoc comparisons, however, there were no significant differences between pilot certificate/rating. A significant main effect occurred for topic on score [Wilks's $\Lambda = 0.46$, $F(3{,}202) = 78.29$, $p < 0.001$, partial $\eta^2 = 0.54$]. Using Bonferroni pairwise post hoc comparisons, regardless of certificate/rating, scores were significantly higher on Data Sources than any other topic in Test 1 ($p < 0.001$), and scores were significantly lower on Storm Definition questions than all other topics ($p < 0.001$, $p = 0.02$, $p < 0.001$). Additionally, scores were significantly higher on Significant Weather questions than Product Attributes ($p = 0.02$). There was no significant interaction between pilot certificate/rating and topic [Wilks's $\Lambda = 0.90$, $F(12{,}534.7) = 1.76$, $p = 0.053$, partial $\eta^2 = 0.03$]. This indicates that the performance trend within the topics of Test 1 was roughly the same for each pilot certificate/rating group.

D. Test 2 Topics Analysis

Test 2 consisted of 25 multiple-choice questions that covered four weather products: METARs (8 questions), PIREPs (6 questions), TAFs (6 questions), and winds aloft (5 questions). The descriptive statistics for Test 2 are shown in Table 4.

A mixed (between and within) analysis of variance was conducted to assess the impact of pilot certificate/rating and product within Test 2 on score. A significant main effect occurred for pilot certificate/rating on Test 2 score [$F(4{,}144) = 4.67$, $p = 0.001$, partial $\eta^2 = 0.12$]. Regardless of product in Test 2, Bonferroni post hoc comparisons revealed that scores for private pilots were significantly lower than ATP ($p = 0.001$) and CFIs ($p = 0.05$). No other significant differences were found. There was a significant main effect for product on Test 2 score [Wilks's $\Lambda = 0.30$, $F(3{,}142) = 110.63$, $p < 0.001$, partial $\eta^2 = 0.70$]. According to the Bonferroni post hoc comparisons, scores were significantly lower on METAR questions than PIREP and winds aloft questions ($p < 0.001$). Scores were significantly higher on PIREP questions than TAF questions ($p < 0.001$). Lastly, scores were significantly higher on winds aloft than all other products ($p < 0.001$). There was no significant interaction found for product and pilot certificate/rating on score [Wilks's $\Lambda = 0.91$, $F(12{,}375.99) = 1.16$, $p = 0.313$, partial $\eta^2 = 0.03$]. This indicates that the test scores within the products in Test 2 were approximately the same for each pilot certificate/rating group.

E. Test 3 Topics Analysis

Test 3 consisted of 23 multiple-choice questions, which covered the products of CIP (5 questions), G-Airmet (13 questions), and GTG (5 questions). The descriptive statistics for Test 3 are shown in Table 5.

A mixed (between and within) analysis of variance was conducted to assess the impact of pilot certificate/rating and product within Test 3 on score. In contrast to Tests 1 and 2, no main effect of pilot certificate/rating occurred for Test 3 score [$F(4{,}145) = 2.25$, $p = 0.59$, partial $\eta^2 = 0.06$]. This indicates that, despite differences in certificate/rating, pilots scored similarly on all the products on Test 3. A significant main effect occurred for product on score for Test 3 [Wilks's $\Lambda = 0.44$, $F(2{,}144) = 90.8$, $p < 0.001$, partial $\eta^2 = 0.56$]. Pairwise comparisons found that scores were significantly higher on GTG interpretation questions than all other products ($p < 0.001$). No other significant differences for products occurred. There was not a significant interaction between product and pilot certificate/rating on Test 3 score [Wilks's $\Lambda = 0.94$, $F(8{,}288) = 1.09$, $p = 0.37$, partial $\eta^2 = 0.03$].

F. Test 4 Topics Analysis

Test 4 consisted of 24 multiple-choice product interpretation questions, which included radar concepts and products (12 questions), SIGMET (7 questions), and thunderstorm concepts (5 questions). Descriptive statistics for Test 4 are shown in Table 6.

A mixed (between–within) analysis of variance was conducted to assess the impact of pilot certificate/rating and product or topic within Test 4 on score. A significant main effect occurred for pilot certificate/rating on score [$F(4{,}193) = 6.16$, $p < 0.001$, partial $\eta^2 = 0.11$]. Bonferroni post hoc tests revealed that scores for private pilots were significantly lower than commercial with instrument, ATP, and CFI pilots ($p < 0.001$). No other significant differences in pilot

Table 3 Test 1: descriptive statistics for score (percentage correct) by topic and pilot certificate/rating

Weather product/topic	Private ($n = 69$), $M(SD)$	Private with instrument ($n = 41$), $M(SD)$	Commercial with instrument ($n = 39$), $M(SD)$	CFI ($n = 35$), $M(SD)$	ATP ($n = 22$), $M(SD)$	Total ($n = 206$), $M(SD)$
Data sources	88.8 (14.8)	94.9 (11.6)	91.2 (16.4)	93.1 (10.8)	96.4 (7.9)	92.0 (13.5)
Significant weather	76.6 (22.3)	82.3 (17.6)	85.1 (19.9)	81.7 (21.3)	86.4 (16.8)	81.2 (20.4)
Storm definition	60.4 (19.1)	71.2 (15.9)	68.5 (21.8)	75.4 (22.3)	78.2 (17.4)	68.9 (20.2)
Product attributes	77.7 (19.7)	77.2 (19.3)	80.0 (16.5)	75.4 (21.7)	72.7 (25.9)	77.1 (20.1)
Total	76.1 (12.6)	81.3 (10.1)	81.2 (11.9)	81.4 (13.2)	83.4 (11.6)	79.5 (11.5)

Table 4 Test 2: descriptive statistics for score by product and pilot certificate/rating

Weather product/topic	Private ($n = 35$), $M(SD)$	Private with instrument ($n = 47$), $M(SD)$	Commercial with instrument ($n = 22$), $M(SD)$	CFI ($n = 21$), $M(SD)$	ATP ($n = 24$), $M(SD)$	Total ($n = 149$), $M(SD)$
METAR	51.1 (19.3)	49.9 (15.5)	49.4 (25.1)	61.3 (12.4)	67.2 (17.2)	54.5 (19.0)
PIREP	72.4 (19.4)	78.9 (15.5)	78.8 (19.4)	80.0 (16.3)	82.6 (18.7)	78.1 (17.8)
TAF	47.1 (20.8)	56.7 (27.7)	56.8 (22.8)	62.4 (21.7)	66.7 (25.1)	56.9 (24.8)
Winds aloft	81.1 (18.8)	85.1 (16.4)	83.6 (14.7)	89.5 (17.5)	90.8 (13.2)	85.5 (16.6)
Total	61.2 (12.8)	65.4 (12.7)	65.1 (16.2)	71.7 (12.3)	75.5 (14.3)	66.0 (14.6)

Table 5 Test 3: descriptive statistics for score (percentage correct) by product and pilot certificate/rating

Weather product/ topic	Private $(n = 40), M(SD)$	Private with instrument $(n = 55), M(SD)$	Commercial with instrument $(n = 11), M(SD)$	CFI $(n = 19), M(SD)$	ATP $(n = 24), M(SD)$	Total $(n = 149), M(SD)$
CIP	63.0 (25.8)	57.1 (23.4)	72.7 (18.5)	67.1 (19.0)	70.6 (24.2)	63.4 (23.8)
G-Airmet	59.2 (17.4)	59.4 (19.4)	69.7 (18.9)	68.4 (16.6)	63.0 (15.3)	61.8 (18.1)
GTG	87.5 (15.5)	90.4 (14.8)	90.9 (13.8)	91.6 (13.9)	93.3 (11.3)	90.3 (14.2)
Total	66.2 (14.4)	65.7 (14.7)	75.0 (15.4)	73.2 (12.3)	71.2 (11.7)	67.7 (14.8)

Table 6 Test 4: descriptive statistics for score by product/topic and pilot certificate/rating

Weather product/topic	Private $(n = 55), M(SD)$	Private with instrument $(n = 46), M(SD)$	Commercial with instrument $(n = 29), M(SD)$	CFI $(n = 22), M(SD)$	ATP $(n = 7), M(SD)$	Total $(n = 159), M(SD)$
Radar	54.0 (16.4)	60.5 (18.3)	66.7 (15.2)	66.5 (19.0)	64.2 (16.8)	60.7 (17.7)
SIGMET	73.5 (14.8)	74.5 (18.8)	83.5 (15.4)	83.9 (18.9)	79.5 (17.5)	77.5 (17.3)
Thunderstorm	56.0 (14.9)	60.0 (15.5)	58.1 (14.7)	60.0 (17.1)	65.3 (16.6)	59.2 (15.7)
Total	60.1 (9.9)	64.5 (12.9)	69.8 (11.3)	70.2 (12.1)	69.0 (12.1)	64.9 (12.3)

Table 7 Test 5: descriptive statistics for score by product and pilot certificate/rating

Weather product/topic	Private $(n = 49), M(SD)$	Private with instrument $(n = 51), M(SD)$	Commercial with instrument $(n = 33), M(SD)$	CFI $(n = 18), M(SD)$	ATP $(n = 23), M(SD)$	Total $(n = 174), M(SD)$
CVA	69.8 (21.7)	77.3 (19.8)	72.9 (23.0)	77.8 (15.2)	80.9 (25.2)	74.9 (21.5)
Satellite	49.6 (29.8)	61.3 (28.9)	59.2 (32.8)	57.1 (31.0)	68.3 (18.8)	58.1 (29.4)
Station plots	37.0 (21.6)	38.9 (21.1)	36.5 (22.1)	38.9 (21.4)	47.1 (20.1)	39.0 (21.4)
Surface prognostic	71.0 (22.4)	74.6 (19.8)	70.6 (21.6)	71.1 (24.9)	67.0 (26.0)	71.5 (22.2)
Total	55.4 (14.8)	61.9 (16.5)	58.8 (19.0)	59.8 (13.3)	65.2 (14.8)	59.4 (16.4)

certificate/rating were found. A significant main effect occurred for product/topic on score [Wilks's $\Lambda = 0.54, F(2{,}192) = 67.69, p < 0.001$, partial $\eta^2 = 0.46$]. Using Bonferroni pairwise comparisons, scores were significantly higher on SIGMET questions than all other topics on Test 4 ($p < 0.001$). There was not a significant interaction between topic/product and pilot certificate/rating [Wilks's $\Lambda = 0.95$, $F(8{,}384) = 1.17, p = 0.32$, partial $\eta^2 = 0.02$].

G. Test 5 Topics Analysis

Test 5 consisted of 23 multiple-choice product interpretation questions, which included CVA (5 questions), satellite (7 questions), station plots (6 questions), and surface prognostic charts (5 questions).

The descriptive statistics for Test 5 are shown in Table 7.

A mixed (between–within) subject analysis of variance was conducted to assess the impact of pilot certificate/rating and product within Test 5 on score. A significant main effect occurred for product on score [Wilks's $\Lambda = 0.37, F(3{,}169) = 96.74, p < 0.001$, partial $\eta^2 = 0.63$]. Bonferroni post hoc comparisons revealed that scores were significantly higher on CVA questions than on satellite and station plot ($p < 0.001$). Scores on station plots were significantly lower than all other products ($p < 0.001$). Surface prognostic scores were higher than satellite scores ($p < 0.001$). There was no significant interaction between pilot certificate/rating and product on score [Wilks's $\Lambda = 0.93, F(12{,}447.4) = 0.996, p = 0.45$, partial $\eta^2 = 0.02$]. This indicates that mean scores on the products within Test 5 were about the same despite the different pilot certificate/ratings.

IV. Conclusions

Increasingly, communicating weather information to pilots occurs primarily through weather products/displays rather than via telephone briefings [24]. Despite sophisticated meteorological systems that provide accurate weather information, recent research indicates that GA pilots have difficulties interpreting the information [15]. A pilot who does not understand the message that aviation weather products convey may be at higher risk of encountering hazardous weather [10]. Interpretation scores in the current study were low to moderate. These results align with prior work [15] and indicate that GA pilots of all certification levels have difficulty interpreting the majority of weather displays and products. These low scores may indicate that the communication of these concepts to GA pilots is insufficient.

The low to moderate interpretability scores are likely attributable to a combination of the complexity of the weather concepts and related technologies as well as the usability of the display design. First, weather concepts themselves (e.g., thunderstorms) as well as the technology behind displays (e.g., radar) are not simple topics. The FAA's primary Advisory Circular (AC 00-06B) that describes weather phenomenon has 22 chapters not including space weather [24]. Additionally, AC 00-45H has six chapters that contain information specifically related to weather products [27]. The length of this report is a direct reflection of the complexity of the topics. Although the inherent complexity of weather and weather technology demands that pilots have a fair amount of knowledge/understanding of the subject matter, one must not forget the role of display design in weather product interpretability [22]. That is, differences in interpretability between the various weather products may also be attributable to the usability (or lack thereof) of the particular display design [28,29]. Effective, human-centered display design is particularly important as information complexity increases, as adding unnecessarily complicated elements in a display for complex information could render the display unusable [30].

In the current study, the lowest interpretability scores included the weather products: station plots, CVA, satellite, and surface prognostic, whereas the highest interpretability scores were for winds aloft, PIREPs, and GTG. Consider one of the low-scoring products: satellite imagery. Among other characteristics, satellite imagery includes the basic categories of infrared, visibility, and water vapor [27]. For correct interpretation, the user must understand the differences between these basic categories as well as the meaning of the images for weather at a particular time and place. This likely means that the user needs both some basic knowledge *and* a user-centered display design that assists the user in avoiding errors and generating correct interpretations. In contrast, PIREPs consist of pilots' descriptions of the weather they encounter during flight, and this represents a much simpler level of complexity. The majority of PIREPs could likely be interpreted with a high level of accuracy by nonpilots and nonweather experts. Thus, one reason for the difference between scores on satellite interpretation and PIREP interpretation could be differences in complexity of the concepts involved and, in turn, the usability of how the information is displayed.

Regarding pilot experience, the current research investigated the effect of pilot certificate/rating on aviation weather product interpretation. Overall, GA private pilots scored significantly lower on all tests compared with all other pilot certificate/rating groups. This finding aligns well with prior research, which indicates that low-hour private pilots incur the majority of weather-related incidences [1]. Based on the current results, private pilots' lack of capability to interpret weather products may be a contributing factor to the weather-related accident rate. Interestingly, although private with instrument pilots scored significantly lower than ATP pilots, on most of the products, no other differences appeared between pilot experience levels. This finding also parallels other research findings that indicate that performance in weather scenarios does not trend with flight hours [31] and may be an indication that existing weather products/displays have *a low level of learnability.*

A few limitations of this study exist. First, the study was a between-groups design in that the participants did not respond to questions on all-weather products. This method was enacted in order to achieve as much participation as possible and to avoid the detracting nature of asking participants to take a 118-question online test. However, a direct comparison on all products was not possible. Another limitation came from the high dropout rate of participants as they proceeded throughout the test. The overall retention rate of about 49% can be concerning due to the possibility of response bias and the results may not be indicative of the true general population. Retention rates could be low due to various factors, including a lack of interest by participants, a lack of time by participants, or the difficulty of the questions. A third limitation was the use of a small number of items to calculate the percentages in the most fine-grained analyses of this dataset.

To avoid weather hazards during flight, pilots need both knowledge of weather concepts and easy-to-interpret weather products. Thus, future research should emphasize both increasing the usability of the weather products as well as improving pilots' weather training. To be clear, while GA pilots need some degree of weather knowledge and skills to read weather displays, it is not necessary, desired, or feasible to achieve an average GA pilot weather knowledge and skill to the level of a meteorologist (e.g., the capacity to interpret more challenging products, such as a Skew-T/Log-P diagram or satellite imagery of water vapor). In contrast, it is feasible to improve aviation weather products/display designs to afford improved interpretability by GA pilots. Improving even one product/display to better convey the hazards to GA pilots could potentially impact thousands of pilots. This would likely be a more efficient strategy to improve flight safety than would a strategy focused on pilot training (e.g., developing training that any one GA pilot may use and which, in turn, may increase his/her understanding of the weather product).

In terms of increasing the usability of weather products, work is needed to implement human factors principles and methods to develop and test GA pilot-centered weather products. As effective display design requires expertise in the particular domain (in this case both aviation and weather) and expertise in the human factors of display design [32], the work on weather display design may be best performed as collaborative research with multidisciplinary teams including industry and government partners. In terms of pilot training, needs include training GA pilots to perform effective self-briefings as well as instructional tools that teach GA pilots to connect what they learn in weather self-briefings with subsequent inflight images. With pilot training playing a key role in improving display interpretation, providing weather training tools and strategies for flight instructors will be key as well.

Acknowledgments

The authors thank the Aviation Weather Branch of the Federal Aviation Administration (FAA) for funding this work (FAA Research Grant 14-G-010). The views expressed in this paper are those of the authors and do not represent the organizations with which they are affiliated or the FAA. The authors express appreciation to the Aircraft Owners and Pilots Association organization for collaborating with us on this study and to Amber Cole, undergraduate research assistant at Embry-Riddle Aeronautical university, for her assistance on this study. The authors thank the journal reviewers who reviewed this paper. Finally, the authors give thanks to all the pilots who participated in this study. It is not easy to answer questions about weather!

References

[1] "27th Joseph T. Nall Report," Air Safety Institute, https://www.aopa.org/-/media/Files/AOPA/Home/Training-and-Safety/Nall-Report/27thNallReport2018.pdf [retrieved April 2021].

[2] "NTSB 2011–2012 Most Wanted List of Transportation Safety Improvements," National Transportation Safety Board, https://www.ntsb.gov/safety/mwl/Pages/mwl_archive.aspx [retrieved April 2021].

[3] "NTSB 2014 Most Wanted List of Transportation Safety Improvements," National Transportation Safety Board, https://www.ntsb.gov/safety/mwl/Pages/mwl_archive.aspx [retrieved April 2021].

[4] "NTSB 2015 Most Wanted List of Transportation Safety Improvements," National Transportation Safety Board, https://www.ntsb.gov/safety/mwl/Pages/mwl_archive.aspx [retrieved April 2021].

[5] "NTSB 2017–2018 Most Wanted List of Transportation Safety Improvements," National Transportation Safety Board, https://www.ntsb.gov/safety/mwl/Pages/default.aspx [retrieved April 2021].

[6] "About Us," Aviation Weather Center, https://aviationweather.gov/info [retrieved April 2021].

[7] Riggio, R., *Introduction to Industrial/Organizational Psychology*, 6th ed., Routledge, New York, 2016, pp. 411–447, Chap. 11.
https://doi.org/10.4324/9781315620589

[8] Fultz, A. J., and Ashley, W. S., "Fatal Weather-Related General Aviation Accidents in the United States," *Physical Geography*, Vol. 37, No. 5, 2016, pp. 291–312.
https://doi.org/10.1080/02723646.2016.1211854

[9] Boyd, D. D., "A Review of General Aviation Safety (1984–2017)," *Aerospace Medicine and Human Performance*, Vol. 88, No. 7, 2017, pp. 657–664.
https://doi.org/10.3357/AMHP.4862.2017

[10] Lanicci, J., Halperin, D., Shappell, S., Hackworth, C., Holcomb, K., Bazargan, M., and Iden, R., "General Aviation Weather Encounter Case Studies," Federal Aviation Administration TR DOT/FAA/AM-12/11, Washington, D.C., 2012.

[11] Parson, S., Beringer, D., Knecht, W., Crognale, M. A., Wiegmann, D., and Beard, B. L., *General Aviation Pilot's Guide to Preflight Weather Planning, Weather Self-Briefings, and Weather Decision Making*, Federal Aviation Administration, Washington, D.C., https://www.faasafety.gov/files/gslac/courses/content/33/346/GA%20Weather%20Decision-Making%20Aug06.pdf [retrieved April 2021].

[12] Burian, B. K., and Jordan, K., "General Aviation Pilot Weather Knowledge and Training: Final Report," Federal Aviation Administration Final Rept. #00-G-020, Washington, D.C., 2002.
https://doi.org/10.1080/24721840.2018.1431780

[13] Latorella, K., and Chamberlain, J., "Graphical Weather Information System Evaluation: Usability, Perceived Utility, and Preferences from General Aviation Pilots," SAE TP 2002-01-1521, Warrendale, PA, 2002.
https://doi.org/10.4271/2002-01-1521

[14] Beringer, D. B., and Ball, J. D., "The Effects of NEXRAD Graphical Data Resolution and Direct Weather Viewing on Pilots' Judgments of Weather Severity and Their Willingness to Continue a Flight," Federal Aviation Administration TR DOT/FAA/AM-04/5, Washington, D.C., 2004.
https://doi.org/10.1177/154193120504900131

[15] Blickensderfer, B. L., Guinn, T. A., Lanicci, J. M., Ortiz, Y., King, J. M., Thomas, R. L., and DeFilippis, N., "Interpretability of Aviation Weather Information Displays for General Aviation," *Aerospace Medicine and Human Performance*, Vol. 91, No. 4, 2020, pp. 318–325.
https://doi.org/10.3357/AMHP.5245.2020

[16] "NEXRAD," National Oceanic and Atmospheric Administration, Washington, D.C., https://www.ncdc.noaa.gov/data-access/radar-data/nexrad [retrieved April 2021].
https://doi.org/10.7289/V5W9574 V

[17] Latorella, K., and Chamberlain, J., “Decision-Making in Flight with Different Convective Weather Information Sources: Preliminary Results,” *Proceedings of the 11th International Symposium for Aviation Psychology*, Dept. of Aviation and Aeronautical Engineering, Ohio State Univ., Columbus, OH, 2001.

[18] Knecht, W. R., “Current NEXRAD Cannot Reliably Enable Safe Flight Around Heavy Weather,” *International Journal of Aviation Psychology*, Vol. 26, Nos. 1–2, 2016, pp. 46–61. https://doi.org/10.1080/10508414.2016.1234936

[19] Cobbett, E. A., Blickensderfer, E. L., and Lanicci, J., “Evaluating an Education/Training Module to Foster Knowledge of Cockpit Weather Technology,” *Aviation, Space, and Environmental Medicine*, Vol. 85, No. 10, 2014, pp. 1019–1025. https://doi.org/10.3357/ASEM.3770.2014

[20] Blickensderfer, E., Lanicci, J., Vincent, M., Thomas, R., Smith, M. J., and Cruit, J., “Training General Aviation Pilots for Convective Weather Situations,” *Aerospace Medicine and Human Performance*, Vol. 86, No. 10, 2015, pp. 881–888. https://doi.org/10.3357/AMHP.4174.2015

[21] Wiggins, M., “Differences in Situation Assessments and Prospective Diagnoses of Simulated Weather Radar Returns Amongst Experienced Pilots,” *International Journal of Industrial Ergonomics*, Vol. 44, No. 1, 2014, pp. 18–23. https://doi.org/10.1016/j.ergon.2013.08.006

[22] Blickensderfer, B., Lanicci, J., Guinn, T., King, J., Ortiz, Y., and Thomas, R., “Assessing General Aviation Pilots’ Understanding of Aviation Weather Products,” *International Journal of Aerospace Psychology*, Vol. 27, Nos. 3–4, 2017, pp. 79–91. https://doi.org/10.1080/24721840.2018.1431780

[23] “Qualtrics,” Software, Ver. 2018, Provo, UT, 2005, http://qualtrics.com.

[24] “Advisory Circular 00-6B,” Federal Aviation Administration, Washington, D.C., https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_00-6B.pdf [retrieved April 2021].

[25] Duke, R., Davis, K., George, T., and Bell, E., *Weather Survey*, Aircraft Owners and Pilots Assoc., Frederick, MD, https://download.aopa.org/advocacy/2019/AOPA_2019_Weather_Survey_Report.pdf [retrieved April 2021].

[26] “IBM SPSS Statistics for Windows,” Ver. 25.0, IBM Corp., Armonk, NY, 2017.

[27] “Advisory Circular 00-45H,” Federal Aviation Administration, Washington, D.C., https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/1030235 [retrieved April 2021].

[28] Remy, B., “Exploring Usability in Web-Based Aviation Weather: An Assessment of the Aviation Weather Center Website,” Ph.D. Dissertation, Florida Inst. of Technology, Melbourne, FL, 2017. https://doi.org/10.1177/1071181319631339

[29] McSorley, J., Kleber, J., and Blickensderfer, B., “Usability Analysis of Aviation Weather Products for General Aviation,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63, No. 1, Nov. 2019, pp. 1903–1907. https://doi.org/10.1177/1071181319631339

[30] Billings, C., *Aviation Automation: The Search for A Human-Centered Approach*, CRC Press, Boca Raton, FL, 2018, p. 370. https://doi.org/10.1201/9781315137995

[31] King, J. M., “An Aviation Weather Preflight Decision Support Tool to Improve GA Pilots Preflight and Inflight Performance,” Ph.D. Dissertation, Embry-Riddle Aeronautical Univ., Daytona Beach, FL, 2020, https://commons.erau.edu/edt/538.

[32] Stanton, N., Salmon, P., Rafferty, L., Walker, G., Baber, C., and Jenkins, D., *Human Factors Methods: A Practical Guide for Engineering and Design*, 2nd ed., CRC Press, Boca Raton, FL, 2017, pp. 1–0, Chap. 1.

P. J. Smith
Associate Editor