Interpretability of Aviation Weather Information Displays for General Aviation

Beth L. Blickensderfer; Thomas A. Guinn; John M. Lanicci; Yolanda Ortiz; Jayde M. King; Robert L. Thomas; Nicholas DeFilippis

BACKGROUND: General Aviation (GA) pilots who encounter hazardous weather inflight have a high probability of incurring fatal accidents. To mitigate this problem, previous research investigated pilot decision making and the effects of new technology. Limited investigations have examined usability and interpretability of observation and forecast weather products available to pilots. Therefore, this study examined the interpretability of weather observation and forecast reports that GA pilots use for preflight weather planning and the impact of pilot certification level on the interpretability of these displays.

METHOD: There were 204 GA pilots (Mean age = 22.50 yr; Median flight hours = 131.0) who completed a 90-item multiple choice Aviation Weather Product Test. The questions portrayed static weather displays available on the NOAA/National Weather Service Aviation Weather Center website. The questions were designed to have high cognitive fidelity in comparison with preflight weather planning tasks.

RESULTS: The results revealed overall low mean interpretability scores (Mean percent correct= 59.29%, SD = 16.01%). The scores for observation products and product attributes were lower for student pilots than experienced pilots. Forecast product scores for student and private pilots did not differ, however, student pilot scores were significantly lower than instrument rated private and commercial pilots.

DISCUSSION: The low interpretability scores indicate that GA pilots misinterpret weather information provided by most weather observation and forecast products. Possible contributing factors to the low product interpretation scores include poor usability and a lack of training. Future research should measure the usability of weather displays designed for pilots.

KEYWORDS: general aviation, information displays, weather, usability.

Blickensderfer BL, Guinn TA, Lanicci JM, Ortiz Y, King JM, Thomas RL, DeFilippis N. *Interpretability of aviation weather information displays for general aviation*. Aerosp Med Hum Perform. 2020; 91(4):318–325.

On a gloomy January day in 2013, a Piper Arrow crashed 2 mi short of the runway at Dover Air Force Base, DE. The pilot had flown 2 h beyond his expected 3-h flight time searching for an airport with satisfactory landing conditions. Although numerous factors contribute to any accident, the National Transportation Safety Board (NTSB) investigation for this tragic flight indicated that the pilot lacked adequate understanding of degrading weather conditions along his flight path and how the conditions could impact his designated destination as well as the surrounding area. Pilots encountering hazardous weather and incurring tragic accidents has become a persistent challenge in United States General Aviation (GA) operations[12,22] (i.e., civil aviation not including scheduled/unscheduled air carriers or commercial space operations).[9] Each year, GA pilots account for 88% of all weather-related aviation accidents in the U.S. and, of these accidents, 61.9% are fatal.[12] From 2003–2007, weather accounted for 733 deaths, or about 146 deaths annually, and the rate has remained relatively constant since 2000.[12] Basic factors underlying the high number of weather-related accidents include that smaller GA aircraft are less resilient to weather hazards, GA flights occur at low altitudes, and weather information is not as accessible to

From the Embry-Riddle Aeronautical University, Daytona Beach, FL, and the University of South Alabama, Mobile, AL.

This manuscript was received for review in August 2018. It was accepted for publication in January 2020.

Address correspondence to: Beth L. Blickensderfer, Human Factors and Systems Department, Embry-Riddle Aeronautical University, 600 S. Clyde Morris Blvd., Dayton Beach, FL 32114; blick488@erau.edu.

DOI: https://doi.org/10.3357/AMHP.5245.2020

GA pilots while inflight as compared to the inflight weather information available to commercial pilots.[18,20] Additionally, human factors researchers have examined potential contributing factors, such as decision-making biases, effects of new technology, and lack of effective training techniques.[1,14,27] An interrelated factor, and the emphasis of the current paper, is the interpretability of weather observation reports and weather forecasts that are used for GA flight planning.

To consider the potential impact that weather may have on their flight, pilots must compile, review, and analyze weather information obtained from a variety of sources. Correct interpretation of weather displays sets the stage for pilots to have an accurate understanding of the weather that may occur during the flight, and in turn, to make appropriate flight plans. GA pilots have access to three major categories of weather information both before and during flight:[21] observation reports, analysis products, and forecasts. Depending on the particular information and the display strategy, the information may be textual, graphic, or oral/verbal. This paper uses the terms "weather displays," "weather information," and "weather reports" interchangeably in reference to all weather products (i.e., the actual products pilots use during preflight weather planning).

First, observation reports describe existing weather conditions (e.g., lightning, precipitation, wind, temperature). A variety of sensors collect the observation data and other systems convey the information to pilots. Examples of observation reports include: Aircraft Reports (AIREPS) and Pilot Reports (PIREPS), Meteorological Terminal Air Report (METARS), Radar Imagery, and Satellite Imagery. Second, analysis products depict enhancements and/or interpretations of observed weather data. Examples of analysis reports include Surface Analysis Charts, Ceiling and Visibility Analysis (CVA), and the Weather Depiction Chart. Third, forecasts are predictions for the development and/or movement of weather conditions. Meteorologists use meteorological observations and mathematical models to develop forecasts. Examples of forecast products available to pilots include Terminal Aerodrome Forecasts (TAFs), Prognostic charts (i.e., "Prog" charts), Area Forecasts, GTGs (graphical turbulence guidance), Convective and Non-Convective SIGMETs (significant meteorological information), Winds Aloft, and Freezing Level.

Numerous sources make the information available to pilots.[11] For example, online tools such as the NOAA/National Weather Service Aviation Weather Center website (hereafter "AWC website") provide an entire suite of weather products/information (https://www.aviationweather.gov/). In addition, the Flight Service Stations (FSSs) provide standard, abbreviated, and outlook weather briefings to pilots. Available via automated telephone recordings and online, the FSS briefings include weather observations, analysis, and forecast information, as well as the opportunity to talk to an FSS specialist and obtain an interpretation of the weather information.[8] During flight, GA pilots may also receive updated weather reports from Enroute Flight Advisory System (EFAS), Hazardous In-flight Weather Advisory System (HIWAS), Air Traffic Control (if workload permits), Automated Weather Observing System (AWOS)/Automated Surface Observation System (ASOS), and the Automated Terminal Information Service (ATIS).

Despite availability of weather information, real-world data and simulation-based research has provided evidence that pilots are not using weather technology effectively and are also susceptible to poor decision making in weather situations.[1] Contributing to these errors and the subsequent possible weather encounters may be pilots' inaccurate and/or incomplete interpretations of the weather information.[18] First, multiple studies have indicated that both GA and commercial pilots fly too close to hazardous weather and seem to ignore FAA recommendations on the distance to maintain from weather hazards such as convective weather.[1,5,28] For example, a study of the en-route portion of GA instrument flight rules (IFR) flights found that a majority of flights violated the FAA recommended separation distance from extreme convection.[5] However, the study could not determine what weather information the pilots of those flights had accessed or attempted to access nor how the pilots interpreted information they did access. Similar results have been replicated in the laboratory when pilots have had full access to weather information. For example, previous research examined the effect that providing GA pilots a variety of weather products consolidated into a portable weather app had on pilots' weather situation awareness.[1] While the study results indicated the technology did have some impact on weather situation awareness, once again, these pilots flew closer than 20 statute miles from hazardous weather cells, thus failing to follow current weather-avoidance recommendations. Again, the results did not include details on pilots' interpretation of the weather. In addition, in a laboratory study with transport pilots, research scientists tested the use of a display which integrated flight trajectories with forecasted weather.[28] The results indicated that if it is relatively easy to achieve a safe distance from weather, the pilots will make an effort to do just that. However, if achieving a safe distance from weather and altering the flight path resulted in increased workload, the pilots often chose to fly closer to the storms instead.[28] It may be that these pilots had not acquired a full understanding of the hazardous conditions and, in turn, chose to fly at unsafe distances.

Another area of research has examined pilot decision making in situations where pilots traveled from Visual Flight Rules (VFR) conditions into Instrument Meteorological Conditions (IMC).[14] Previous research indicates that, between the years 1982 and 2013, an average of 124 fatal weather-related accidents occurred annually, and, of those, 71% (5681 total for the period) were related to flying into IMC.[13] Research suggests possible causes for this decision to enter IMC from VFR conditions flight include unrecognized changes in the weather, incorrect understanding of the severity of expected weather, improper risk assessment, and decision biases (e.g., framing and anchoring) that adversely impacted their decisions.[14,15,27] All of these factors may be exacerbated by difficult-to-interpret weather displays and products, which in turn, limit the utility of the information for pilot decision making.

Piloting an aircraft requires a variety of higher-order cognitive skills, and these include interpreting aviation weather information and forecasts and applying the information correctly to flight. If pilots are unable to interpret weather products and relate the information to flight and/or out-the-window weather cues, it will be impossible for them to perform effectively in situations in which they encounter weather and will leave them susceptible to decision making errors.[17,27] Most existing research that examined GA pilots' use of weather technology and subsequent decision making has not included direct assessments of the accuracy of the pilots' weather interpretation. Thus, while these studies demonstrated that pilots make errors, the reason(s) for the errors remain somewhat unclear, and poor display interpretability cannot be ruled out as a contributing factor.

Limited research has examined interpretability of individual weather products and displays in GA contexts. The research that has examined weather product usability tends to focus on graphical vs. textual displays; including Next Generation Radar displays (i.e., NEXRAD).[2,5,19] In an early study of NEXRAD with simulated flights, results indicated that GA pilots flew closer to storms when viewing high resolution radar displays than with low resolution displays or no radar displays.[2] These results may indicate the pilots were misinterpreting the information presented in the high resolution displays. Additionally, two studies have provided a direct assessment of knowledge of thunderstorm concepts as well as radar interpretation, and both studies revealed that pilots had difficulty understanding this information.[3,6] These and other studies indicate limitations with pilots' interpretation of NEXRAD.

In addition, further research conducted usability testing on graphical weather information systems including the early NEXRAD displays.[19] In a study with 12 instrument-rated GA pilots, the results showed that pilots embraced the graphical products and had increased confidence with the products, but the study also found some limitations in the degree to which pilots could interpret the displays accurately.[19] Other research has focused on transport pilots. In this particular study, research scientists observed two-person airline pilot crews' use of electronic flight bags (EFB) which included some weather information.[7] Basic usability issues were described, and the researchers also observed that pilots performed better when weather information was presented graphically. Perhaps the most detailed examination performed on pilots' behavior related to usability is a study evaluating the design of graphical weather products.[23,24] In comparing a commercially available graphical display, an ergonomically redesigned display, and an ordinary text statement, results indicated that participants recalled more information from a graphical display designed with the ergonomic principles as compared to existing graphical displays and textual formats of the same data.[23,24]

While some research has indicated a need for improved aviation weather product usability, no research exists on the breadth of the problem. The existing research has focused primarily on a small number of products and has used small sample sizes which do not cover the breadth of GA pilots' level of certification and/or rating. This leaves research gaps regarding the interpretability of all aviation weather products by GA pilots. Thus, the purpose of this study was to provide a broad assessment of the interpretability of weather observation reports and weather forecasts used for GA preflight weather planning and examine the impact of pilot certification level on the interpretability of these displays.

METHODS

Subjects

Participants in the study ($N = 204$) were recruited from a southeastern U.S. university and from a midwestern Air Venture airshow. Participant age ranged from 15 to 66 (Mean age = 22.50 yr, SD = 7.60). Participants were divided between four categories of FAA GA flight certificate/rating, which were: Student Pilot ($N = 41$, Mean flight hours = 38.37 h, SD = 30.83; Median = 35.00), Private Pilot ($N = 72$, Mean flight hours = 128.77 h, SD = 118.50; Median = 105.00), Private Pilot with Instrument Rating ($N = 50$, Mean flight hours = 211.46, SD = 196.68; Median = 172.00), and Commercial Pilot with Instrument Rating ($N = 41$, Mean flight hours = 479.87, SD = 1015.22; Median = 260.00), resulting in 204 participants in total (Mean flight hours = 201.43, SD = 490.36; Median = 131.00). The study was approved in advance by the Embry-Riddle Aeronautical Institutional Review Board for the protection of human participants, and all participants signed an informed consent form. For their participation, each pilot at the university was given \$20 plus \$0.31 per question answered correctly, while each pilot at the airshow was given a \$100 gift card.

Equipment

The materials consisted of a demographic questionnaire and the Aviation Weather Product Test. The demographic questionnaire consisted of 33 items which were designed to obtain basic information about participants such as age, flight training, flight hours, and meteorological training (e.g., when and where they received weather training/course and how frequently they use aviation weather products). The demographic questionnaire was implemented using an online survey website (surveymonkey.com).

The study used an Aviation Weather Product Test that was previously developed.[4] The test evaluates the interpretability of aviation weather products used for GA flight planning, and the questions focus primarily on the graphical or textual weather products that are hosted on the AWC website. The original test consisted of 95 multiple-choice questions and each had 3–4 answer options (a, b, c or a, b, c, d) with only one correct answer.[4] The questions emphasized application of information. The test required respondents to interpret the weather products, just as they would in actual flight planning. Thus, the test achieved a high level of cognitive fidelity.[4] Five questions from the Aviation Weather Product Test were not included in the analyses: two questions relating to storm definitions rather than display interpretation, one question on station plots, and two questions that had awkward wording were removed from the data.[4]

The topics included in the current study were weather observation products, weather forecast products, and product attributes. While a full description of these products is beyond the scope of this paper, interested readers can find them in the FAA Advisory Circular 00-45H.[11] Briefly, the weather observation products report current weather conditions (e.g., precipitation, clouds, fog, wind speed/direction, etc.). The observation products include Meteorological Report (METAR), Pilot Report (PIREP), Radar, Satellite, Ceiling and Visibility Analysis, and Winds Aloft).

See below an example of an observation product question:

Example 1: In the METAR remarks for your destination airport you notice the comment "CB DSNT N MOV N." Based on this information, which of the following is true?

a) Cumulonimbus clouds are north but within 5 nautical miles of the airport and approaching the airport.
b) Cumulonimbus clouds are less than 10 statute miles north of the airport and approaching the airport.
c) Cumulonimbus clouds are less than 10 statute miles north of the airport and moving away from the airport.
d) Cumulonimbus clouds are more than 10 statute miles north of the airport and moving away from the airport.[4]

The weather forecast products display weather conditions that are predicted to occur in the future. The forecast products include Low-level Significant Weather Chart (LL SigWx), Graphical-Airman's Meteorological Product (G-AIRMET), Graphical Turbulence Guidance (GTG), Terminal Area Forecast (TAF), Surface Prognostic Chart (Surface Prog), Significant Meteorological Information (SIGMET), National Convective Weather Forecast Product (NCWF), and Current and Forecast Icing Product (CIP/FIP).

See below an example of a forecast product question:

Example 2: Referring to the below TAF, the prevailing conditions starting on the 11th of the month at 08Z are considered valid until what time?

KPIA 101745Z 1018/1118 18016G25KT 6SM -RA BR
SCT020 OVC040
TEMPO 1018/1021 2SM TSRA BR BKN015CB
FM102100 20014G23KT 4SM -RA BR SCT007 OVC015
TEMPO 1021/1024 1 1/2SM SHRA BR BKN007
FM110100 23010KT 4SM -RA BR SCT006 OVC012
FM110400 35006KT 6SM BR VCSH SCT010 OVC020
FM110800 34003KT P6SM BKN050
TEMPO 1112/1115 BKN030

a) The 11th of the month at 1200Z
b) The 11th of the month at 1100Z
c) The 11th of the month at 1800Zs
d) Unknown or unspecified.[4]

The product attributes questions were topics key to correct use and interpretation applicable to both weather observation and weather forecast products. These questions focused on valid sources of weather information (i.e., source validity) and product limitations. Source validity questions measured participants' interpretation of product source limitations, such as recognizing FAA-approved sources of information and the availability of certain data sources during different phases of flight, such as preflight and inflight (e.g., AWC website, 1-800WxBrief). Similarly, product limitations questions covered topics such as product issuance times, valid times, and product reliability.

See below an example of a product attributes question:

Example 3: Why is it unsafe for a pilot to fly through a gap between thunderstorms as shown on a real-time cockpit display of ground-based radar data?

a) Ground-based radar data cockpit displays are not capable of accurately depicting strong thunderstorms
b) Ground-based radar cockpit displays only show recent thunderstorm activity, not current storm activity
c) Gaps in ground-based radar echoes often indicate embedded thunderstorms are occurring at that location.
d) Gaps in ground-based radars are indicative of beam attenuation and possible IFR conditions within the gap.[4]

Percentage correct scores were calculated for each participant. The overall interpretation score was the percentage correct across all 90 items on the Aviation Weather Product Test. The Cronbach's alpha was 0.92; this high alpha score indicates a strong degree of intercorrelation among the questions. An additional percentage correct score was calculated for each of the three respective categories (see **Table I**): Observation products (34 items; $\alpha = 0.80$), Forecast products (48 items; $\alpha = 0.88$) and Product attributes (8 items; $\alpha = 0.59$). In addition, subscores were also calculated. The subscores were the percentage correct scores for each respective observation product (METAR, PIREP, Radar, Satellite, Ceiling and Visibility Analysis, and Winds Aloft) (see **Table II**); percentage correct for each respective forecast product (Low-level Significant Weather Chart, G-AIRMET, GTG, TAF, Surface Prog, SIGMET, NCWF, CIP/FIP) (see **Table III**); and for each respective Product attributes topic (source validity and product limitations) (see **Table IV**).

Procedure

Upon arriving at the testing site, each participant was given the informed consent form to review and sign. Next, the university participants completed the Demographic Questionnaire and the Aviation Weather Product Test on a Dell desktop computer in a secure testing location on the campus. For the participants recruited at the airshow, the participants completed the Demographic Questionnaire online, using their own personal devices, but they completed the Aviation Weather Test by reading a hardcopy of the questions and recording responses on a paper score sheet. No time restriction was used; all participants completed the test at their own pace. Upon completion of the test, each participant was debriefed and given their compensation.

Statistical Analyses

To assess the differences in product usability and interpretability, a series of analyses were conducted using the Statistical Package for the Social Sciences (SPSS).[16] First, analyses were

Table I. Overall Weather Interpretation Scores (Percentage Correct).

	ITEMS	ALPHA	STUDENT M (SD) $N = 41$	PRIVATE M (SD) $N = 71$	PRIVATE W/INSTRUMENT M (SD) $N = 50$	COMMERCIAL W/INSTRUMENT M (SD) $N = 41$	INTERPRETATION SCORE (TOTAL) M (SD) $N = 204$
Observation Interpretation	34	0.80	49.13 (15.21)	58.49 (15.38)	63.06 (15.59)	67.14 (15.23)	59.47 (16.61)
Forecast Interpretation	48	0.88	48.27 (15.20)	56.35 (18.45)	62.21 (14.61)	66.41 (16.80)	58.18 (17.62)
Product Attributes	8	0.59	55.18 (19.56)	68.58 (20.87)	76.75 (16.56)	77.44 (20.58)	69.67 (21.10)
Total	90	0.92	48.92 (14.26)	57.78 (15.84)	63.29 (13.53)	67.43 (15.01)	59.29 (16.01)

N = number of participants; M = mean; SD = standard deviation; Items = number of items; Alpha = Cronbach's alpha.

conducted across all responses to examine overall results. Next, separate analyses were performed for each respective area: observation products, forecast products, and product attributes. Greenhouse Geiser *F*-values were used for all analyses.

RESULTS

First, a 4 × 3 mixed analysis of variance (ANOVA) was run to assess the impact of pilot certification/rating (between factor) and weather product category (within factor) on overall interpretation score. Table I displays the descriptive statistics for pilot certificate/rating (Student, Private, Private with Instrument, and Commercial with Instrument) and weather product category (Forecast Interpretation, Observation Interpretation, and Product Attributes.

There was no significant interaction observed between pilot certification/rating and weather product category [F (4.76, 317.53) = 1.29, $P = 0.27$, partial $\eta^2 = 0.02$]. A significant main effect of pilot certificate/rating on overall interpretation scores was observed [F (3, 200) = 14.14, $P < 0.001$, partial $\eta^2 = 0.18$], with 18% of variance in total score accounted for by pilot certification/rating. Bonferroni post hoc tests revealed that, regardless of the weather product category, interpretation scores were significantly lower from student pilots than any other pilot certification/rating ($P = 0.003$), while interpretation scores from commercial with instrument rated pilots were significantly higher than those from private pilots ($P = 0.011$). No other significant differences occurred ($P > 0.05$).

A main effect also occurred for weather product category on overall interpretation score [F (1.59, 317.53) = 61.39, $P < 0.001$, partial $\eta^2 = 0.24$], with 24% of variance in total score accounted for by product category. Post hoc paired samples *t*-tests indicated that, regardless of pilot experience level, the Product attributes scores were significantly higher than both the Observation scores ($P < 0.001$) and the Forecast scores ($P < 0.001$). No significant difference occurred between Observation scores and Forecast scores ($P = 0.43$).

The second analysis was a 4 × 6 mixed ANOVA to determine the impact of pilot certification/rating (Student, Private, Private with Instrument, and Commercial with Instrument) and observation product (METAR, PIREPs, Radar, Satellite, CVA and Winds Aloft) on observation interpretation scores. The means of these subcategories are found in Table II.

No significant interaction occurred between pilot certification/rating and observation product [F (10.90, 726.35) = 0.92, $P = 0.52$, partial $\eta^2 = 0.01$]. There was a significant main effect for pilot certification/rating [F (3, 200) = 9.74, $P < 0.001$, partial $\eta^2 = 0.13$]. Bonferroni post hoc tests indicated observation interpretation was significantly lower for student pilots than for the other pilot certification/rating groups ($P = 0.005$). No other significant differences occurred ($P > 0.05$).

A significant main effect occurred for observation product [F (3.63, 726.35) = 79.86, $P < 0.001$, partial $\eta^2 = 0.29$], indicating that the observation product accounted for 29% of the variance in scores. Differences within the observation product subcategories were found using Bonferroni post hoc paired samples *t*-tests. Regardless of pilot certificate/rating, there was a significant difference between all observation products ($P < 0.05$) except between METAR and CVA ($P = 0.10$), RADAR and CVA ($P = 0.47$), and Satellite and CVA ($P = 1.0$). Winds Aloft had higher interpretability than any other observation

Table II. Observation Product Interpretation Scores (Percentage Correct).

	ITEMS	ALPHA	STUDENT M (SD) $N = 41$	PRIVATE M (SD) $N = 71$	PRIVATE W/INSTRUMENT M (SD) $N = 50$	COMMERCIAL W/INSTRUMENT M (SD) $N = 41$	OBSERVATION PRODUCT SCORE (TOTAL) M (SD) $N = 204$
Winds Aloft	3	0.54	73.17 (33.52)	86.57 (20.68)	86.00 (25.28)	91.06 (18.29)	84.64 (25.06)
PIREPs	4	0.53	63.21 (32.11)	76.39 (24.43)	85.00 (25.25)	85.37 (20.14)	77.66 (26.72)
Radar	10	0.53	51.22 (20.76)	58.19 (21.84)	66.00 (18.84)	64.88 (17.62)	60.05 (20.71)
CVA	2	0.37	43.90 (35.70)	58.33 (38.46)	52.00 (40.36)	62.20 (39.97)	54.66 (38.96)
Satellite	7	0.66	41.35 (27.69)	52.78 (27.82)	58.00 (25.06)	64.11 (26.75)	54.04 (27.78)
METAR	8	0.45	38.41 (20.04)	44.44 (18.87)	47.00 (19.00)	55.79 (20.94)	46.14 (20.23)
Observation Product Total	34	0.80	49.13 (15.21)	58.49 (15.38)	63.06 (15.59)	67.14 (15.23)	59.47 (16.61)

Note: PIREP, Pilot Report; CVA, Ceiling and Visibility Analysis; METAR, Meteorological Terminal Aviation Routine Weather Report.
The product Winds Aloft is considered an Analysis/Forecast Product by FAA; however, because it provides current as well as forecast wind information, Winds Aloft was analyzed with Observation products.
N = number of participants; M = mean; SD = standard deviation; Items = number of items; Alpha = Cronbach's alpha.

Table III. Forecast Product Interpretation Scores (Percentage Correct).

	ITEMS	ALPHA	STUDENT M (SD) $N = 41$	PRIVATE M (SD) $N = 71$	PRIVATE W/INSTRUMENT M (SD) $N = 50$	COMMERCIAL W/INSTRUMENT M (SD) $N = 41$	FORECAST PRODUCT SCORE (TOTAL) M (SD) $N = 204$
GTG	2	0.36	79.27 (31.58)	79.86 (32.13)	84.00 (31.04)	81.71 (26.82)	81.13 (30.58)
LLSigWx	5	0.43	65.85 (26.92)	70.56 (24.89)	78.40 (21.70)	78.05 (19.39)	73.04 (23.92)
Surface Prog	5	0.59	63.54 (30.13)	68.26 (27.40)	76.40 (23.45)	75.61 (27.39)	70.78 (27.34)
SIGMETs	11	0.65	47.67 (19.37)	60.98 (21.82)	64.91 (18.55)	72.73 (18.63)	61.63 (21.42)
CIP/FIP	4	0.31	40.85 (25.47)	52.78 (28.01)	57.50 (24.87)	59.15 (30.49)	52.82 (27.88)
TAF	6	0.55	38.21 (23.93)	49.31 (26.45)	51.33 (22.04)	61.38 (26.47)	50.00 (25.84)
G-AIRMET	12	0.64	39.63 (18.52)	45.49 (21.85)	53.67 (16.16)	57.93 (21.16)	48.82 (20.72)
NCWF	3	0.11	39.84 (26.06)	42.59 (31.76)	49.33 (27.96)	52.03 (25.87)	45.59 (28.79)
Total	48	0.88	48.27 (15.20)	56.35 (18.45)	62.21 (14.61)	66.41 (16.80)	58.18 (17.62)

Note. GTG, Graphical Turbulence Guidance; LLSigWX, Low Level Significant Weather Chart; Surface Prog, Surface Prognostic Chart; SIGMET, Significant Meteorological Information; CIP/FIP, Current Icing Product/Forecast Icing Product; TAF, Terminal Aerodrome Forecast; G-AIRMET, Graphical Airmen's Meteorological Information; NCWF, National Convective Weather Forecast. M = mean; SD = standard deviation; Items = number of items; Alpha = Cronbach's alpha.

product ($P = 0.025$), while METARs had the lowest interpretability scores ($P = 0.003$). PIREPs had the second highest interpretability scores ($P = 0.025$), followed by RADAR as the third highest ($P = 0.029$).

Next, a 4 × 8 mixed ANOVA examined the impact of pilot certificate/rating (Student, Private, Private with Instrument and Commercial with Instrument) and Forecast products (LLSigWx, G-AIRMET, GTG, TAF, Surface Prog Chart, SIGMETs, NCWF, and CIP/FIP) on interpretation score. The means are found in Table III.

No significant interaction occurred between pilot certification/rating and forecast product [F (17.35, 1156.37) = 1.10, $P = 0.34$, partial $\eta^2 = 0.02$]. A significant main effect was observed for pilot certification/rating [F (3, 200) = 6.94, $P < 0.001$, partial $\eta^2 = 0.09$], in which 9% of the variance in forecast interpretation score was related to pilot certification/rating. Bonferroni post hoc comparisons revealed that forecast products had significantly lower interpretability from student pilots than private with instrument ($P = 0.003$) and commercial with instrument pilots ($P < 0.001$). No other significant differences occurred ($P > 0.05$).

A significant main effect occurred for forecast product [F (5.78, 1156.37) = 80.52, $P < 0.001$, partial $\eta^2 = 0.29$], indicating the forecast product accounts for 29% of the variance in scores. Post hoc paired samples *t*-tests indicated that, regardless of pilot certificate/rating, the GTG interpretation scores were significantly higher than any other forecast product ($P = 0.009$). The second highest scores were LL SigWx which were higher than all the remaining forecast product ($P = 0.009$) except Surface Prog Charts where no significant difference occurred ($P = 1.0$). The third highest scores occurred for Surface Prog, and scores on Surface Progs were significantly higher than G-AIRMET, TAF, SIGMET, NCWF, and CIP/FIP items ($P < 0.001$). SIGMET scores were significantly lower than LL SigWx, GTG, and Surface Prog Charts ($P < 0.001$), but were significantly higher than G-AIRMET, TAFs, NCWF, and CIP/FIP ($P < 0.001$). No significant differences occurred between G-AIRMETs and TAFs ($P = 1.0$), G-AIRMETs and NCWF ($P = 1.0$), G-AIRMETs and CIP/FIP ($P = 1.0$), and NCWF and CIP/FIP ($P = 1.0$). No significant differences also occurred between TAF scores and NCWF ($P = 1.0$) and CIP/FIP ($P = 1.0$) scores.

The final analysis was a 4 × 2 mixed ANOVA to determine the impact of pilot certificate/rating (Student, Private, Private with Instrument, and Commercial with Instrument) and the Product attributes (Product limitations and Source validity) on score. The descriptive statistics are found in Table IV.

No significant interactions occurred between the pilot certification/rating and product attributes [F (3, 200) = 0.33, $P = 0.80$, partial $\eta^2 = 0.01$]. There was a significant main effect observed for pilot certificate/rating [F (3, 200) = 11.3, $P < 0.001$, partial $\eta^2 = 0.15$], where 15% of the variance in score was related to pilot certification/rating. Bonferroni post hoc comparisons revealed that product attributes scores were significantly lower for student pilots than all other pilots ($P = 0.004$). No other significant differences occurred ($P > 0.05$).

A significant main effect for the subcategories within product attributes on score also occurred [F (1, 200) = 5.78, $P = 0.02$, partial $\eta^2 = 0.03$], indicating that the subcategories within product attributes accounted for 3% of variance in scores. Source validity questions had higher scores than did product limitation items.

Table IV. Product Attributes Scores (Percentage Correct).

	ITEMS	ALPHA	STUDENT M (SD) $N = 41$	PRIVATE M (SD) $N = 71$	PRIVATE W/INSTRUMENT M (SD) $N = 50$	COMMERCIAL W/INSTRUMENT M (SD) $N = 41$	PRODUCT ATTRIBUTES SCORE (TOTAL) M (SD) $N = 204$
Product Limitations	5	0.42	54.15 (22.47)	66.39 (20.65)	73.60 (20.00)	76.10 (21.55)	67.65 (22.29)
Source Validity	3	0.48	56.91 (28.13)	72.22 (33.10)	82.00 (24.48)	79.67 (26.75)	73.04 (30.10)
Total	8	0.59	55.18 (19.56)	68.58 (20.87)	76.75 (16.56)	77.44 (20.58)	69.67 (21.10)

M = mean; SD = standard deviation; Items = number of items; Alpha = Cronbach's alpha.

DISCUSSION

Prior research has indicated that weather products and technology, such as radar displays, do not necessarily help GA pilots avoid hazardous weather conditions.[1,5,15] Flight related weather information displays may be difficult to interpret, and GA pilots may not easily discern the implications of that information for flight.[23] Results of the current study indicate that a gap exists in the interpretability of weather observation and forecast products used by GA pilots. The low weather product interpretation scores may indicate unusable displays, inadequate pilot training, or both.

In a close inspection of a series of weather related incidents, previous research indicates that the weather observations and forecasts were available and accurate, and yet pilots still flew into hazards.[18] Previous research also noted that the pilots had a lack of appreciation for what the weather products conveyed, and attributed this to a lack of correct interpretation of the weather products and, in turn, misunderstanding the implications for flight.[18] The results of the current study are in alignment with this finding.

While interpretability scores were lower for student pilots than for experienced pilots on observation and product attributes, the forecast results differed slightly. For forecast, the interpretability for student pilots was similar to that for private pilots—which indicates no impact of the private certification. However, improvement did appear when comparing student pilots to pilots who had achieved instrument ratings or commercial certificates. Thus, although additional flight certificate/ratings indicate that a pilot has achieved higher flight knowledge and skills, additional pilot training did not relate to improved interpretation of observation products but it did for forecast products. These results partially echo previous findings that flight hours are not related to improved performance in unexpected weather situations.[26] Finally, the product attributes questions yielded higher scores than the weather observations and weather forecast questions. The higher scores on source validity questions in comparison to product limitation questions may indicate that while pilots know where to find the weather information, they cannot determine limitations of the information. Taken together, these results may indicate a low usability of many weather observation and forecast products.

When looking within weather observation data, the highest scores were on Winds Aloft and PIREPS while traditional text based METARS had the lowest scores. This could be due to a variety of underlying factors. For instance, pilots may use decoded METARS more often than the traditional coded format which was used in this study. Further, the traditional coded format has been criticized for a lack of adherence to known principles of cognitive memory and recall. Regarding forecasts, the highest scoring category was GTG and the lowest scoring was the G-AIRMET, TAF, NCWF and CIP/FIP forecasts. These results may be surprising, considering that the GTG product was recently introduced to the GA community and pilots have had less time to use it as compared to other forecast products. However, from a display design and usability perspective, the results are less surprising. Specifically, the GTG is a color-coded graphical product that also features a simple legend. Furthermore, this product avoids information overload by only displaying turbulence information, compared to G-AIRMETS, which can indicate multiple phenomena, such as ceilings, turbulence, and icing.

Although the CIP/FIP, GTG, and NCWF are graphical products, pilots may not use them very often. This may be because prior to the issuance of FAA Advisory Circular 00-45G, Change 2, these products were considered supplementary rather than primary products which may have limited their use.[10] In addition, considering the pilot participants were mostly from the southeastern region, they may be less concerned about icing conditions than other pilot populations. This factor may affect how often this sample of pilots used the CIP/FIP and other icing weather products.

Regarding interpreting radar displays, the results in this study parallel other research indicating that pilots struggle to interpret radar correctly.[2,3] Despite the graphical nature, the intricacies of the technology are not grasped by the GA users.[6]

This study was conducted with relatively young, low hour pilots, many of whom fly extensively in the southeastern U.S. where freezing temperatures are infrequent. Research is needed on a sample more generalizable to overall general aviation pilots in the U.S. in terms of older pilot age, greater flight hours, and varying geographic regions of residence. Furthermore, while the exam consisted of 90 total items, some analyses included a relatively small number of items per product. Also, the product interpretation exam did not include two recently introduced weather products: the Graphical Forecasts for Aviation (GFA) and the Aviation Forecasts. Furthermore, although this written test was able to examine basic weather product interpretability, these scores do not reflect pilots' capability to interpret and apply weather product information to an authentic, planned flight.

In considering interpretability of weather information, one must consider the influence of technology on human performance. Although technology is commonly viewed as a "way out" of the human error problem, technology itself can generate serious human performance decrements, and this may be occurring with respect to GA weather-related accidents. In the past two decades, an influx of weather technology has become available to pilots prior to and during flight, with aims that the new technology would reduce the weather accident rate.[21] Today, ample research exists demonstrating that NEXRAD technology, in particular, is not necessarily helping GA pilots to avoid hazardous weather.[2,28] The current study further supports this notion, as even the more experienced pilots struggled to interpret weather products–including radar images. The results of this study can be taken as indication of possible design weaknesses regarding the usability of these weather products by GA pilots.

Future research should include the usability of weather displays designed specifically for pilots. Until more usable weather products are available, however, the complex weather technologies necessitate additional training for GA pilots. The most effective training would occur in a flight specific context with ample opportunities to practice applying the concepts.[17] Fortunately, new, less expensive simulation capabilities are becoming available, as adequate simulations are essential for pilots to practice interpreting and applying weather observations and forecasts to flight.[17,25] For GA pilots who do not have access to simulation-based training, tools such as video presentations could provide demonstrations and insights regarding correct weather product interpretation.

ACKNOWLEDGMENTS

This research was sponsored by the FAA Weather Technology-in-the-Cockpit (WTIC) Program and was conducted under FAA #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 wish to thank all of the pilots who participated in this study.

Financial Disclosure Statement: The authors have no competing interests to declare.

Authors and affiliations: Beth L. Blickensderfer, Ph.D., Thomas A. Guinn, Ph.D., Yolanda Ortiz, M.S., Jayde M. King, M.S., Robert L. Thomas, Ph.D, and Nicholas DeFilippis, M.S., Embry-Riddle Aeronautical University, Daytona Beach, FL; John M. Lanicci, Ph.D., University of South Alabama, Mobile, AL.

REFERENCES

1. Ahlstrom U, Ohneiser O, Caddigan E. Portable weather applications for general aviation pilots. Hum Factors. 2016; 58(6):864–885.
2. Beringer DB, Ball JD. The effects of NexRad graphical data resolution and direct weather viewing on pilots' judgments of weather severity and their willingness to continue a flight. Washington (DC): Federal Aviation Administration, Office of Aerospace Medicine; 2004 Mar. Report No: DOT/FAA/AM-04/5[Accessed 28 Jan 2020.] Available from: https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2000s/media/0405.pdf
3. Blickensderfer B, Lanicci J, Guinn T, Thomas R, King J, et al. Assessing general aviation pilots understanding of aviation weather products. Intl J Aerosp Psychol. 2017; 27(3-4):79–91.
4. Blickensderfer EL, Lanicci JM, Vincent MJ, Thomas RL, Smith MJ, Cruit JK. Training general aviation pilots for convective weather situations. Aerosp Med Hum Perform. 2015; 86(10):881–888.
5. Boyd DD. In-flight decision-making by general aviation pilots operating in areas of extreme thunderstorms. Aerosp Med Hum Perform. 2017; 88(12):1066–1072.
6. Cobbett EA, Blickensderfer EL, Lanicci JM. Evaluation an Education/Training Module to Foster Knowledge of Cockpit Weather Technology. Aviat Space Environ Med. 2014; 85(10):1019–1025.
7. Evans ET, Young SD, Daniels TS, Meyer RR. Usability of EFBs for viewing NOTAMs and AIS/MET data link messages. Proceedings of the IEEE/AIAA 32nd Digital Avionics Systems Conference; 2013 Oct 6-10, East Syracuse, NY. Piscataway (NJ): IEEE; 2017.
8. Federal Aviation Administration. Aeronautical Information Manual Official Guide to Basic Flight Information and ATC Procedures, 3rd rev. Newcastle (WA): Aviation Supplies and Academics, Inc; c2019. Chapter 7, Safety of Flight; p.485.
9. Federal Aviation Administration. Aeronautical Information Manual Official Guide to Basic Flight Information and ATC Procedures, 3rd rev. Newcastle (WA): Aviation Supplies and Academics, Inc; c2019. Glossary, G; p.685.
10. Federal Aviation Administration. Aviation Weather Services. Advisory Circular 00-45G, Change 2. [Cited 2014] Available from: https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_00-45G_CHG_1-2.pdf
11. Federal Aviation Administration. Aviation Weather Services. Advisory Circular 00-45H, Change 1. [Cited 2018] Available from: https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_00-45H_Chg_1.pdf
12. Federal Aviation Administration. Weather-related aviation accident study 2003–2007. [Cited 2010] Available from: https://www.asias.faa.gov/i/studies/2003-2007weatherrelatedaviationaccidentstudy.pdf
13. Fultz AJ, Ashley WS. Fatal weather-related general aviation accidents in the United States. Physical Geography. 2016; 3646:1–22.
14. Goh J, Wiegmann DA. Human factors analysis of accidents involving visual flight rules flight into adverse weather. Aviat Space Environ Med. 2002; 73:817–822.
15. Hunter DR. Risk perception and risk tolerance in aircraft pilots. Washington (DC): Federal Aviation Administration; 2002 Sept. (Rep. No. DOT/FAA/AM-02/17).
16. International Business Machines Corporation. Statistical Package for Social Sciences Statistics for Windows. Version 24.0 [software]. 2016 April 9 [cited 2019 Sept 5]. Available from: https://www.ibm.com/analytics/spss-statistics-software.
17. Johnson CM, Wiegmann DA. VFR into IMC: Using simulation to improve weather-related decision-making. Int J Aviat Psychol. 2015; 25(2):63–76.
18. Lanicci J, Halperin D, Shappell S, Hackworth C, Holcomb K, et al. General Aviation Weather Encounter Case Studies. Oklahoma City (OK): FAA Office of Aerospace Medicine, Civil Aerospace Medical Institute; 2012; Report DOT/FAA/AM-12/11.
19. Latorella KA, Chamberlain JP. Graphical weather information system evaluation: Usability, perceived utility, and preferences from general aviation pilots. Hampton (VA): Crew/Vehicle Integration Branch, NASA Langley Research Center; 2002; Report No: 2002-01-1521.
20. Latorella KA, Chamberlain JP. Tactical vs. strategic behavior: General aviation piloting in convective weather scenarios. Proceedings of the HFES 44th Annual Meeting; 2002 Sept; Baltimore, MD. Santa Monica (CA): Human Factors and Ergonomics Society; 2002
21. National Oceanic and Atmospheric Administration & Federal Aviation Administration. Aviation weather services, AC 00-45G change 2. [Cited 2014] Available from: https://www.faa.gov/documentlibrary/media/advisory_circular/ac_00-45g_chg_1-2.pdf
22. National Transportation Safety Board (NTSB). NTSB 2014-2015 Most wanted list of Transportation Safety Improvements: General Aviation: Identify and Communicate Hazardous weather. [Cited 2014] Available from: https://www.ntsb.gov/safety/mwl/Pages/mwl7_2014.aspx
23. O'Hare D, Stenhouse N. Redesigning a graphic weather display for pilots. Ergon Des. 2008; 16:11–15.
24. O'Hare D, Stenhouse N. Under the weather: An evaluation of different modes of presenting meteorological information for pilots. Appl Ergon. 2009; 40(4):688–693.
25. Ortiz Y, Blickensderfer B, Thomas R, Bois R, Allen M, Mikail S. Flight simulation without an instructor: Teaching missed approaches. Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), 2016 Nov 28-Dec 01, Orlando, FL. Arlington (VA): National Training and Simulation Association.
26. Wiggins M, O'Hare D. Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. J Exp Psychol Appl. 1995; 1(4):305–320.
27. Wiggins M, O'Hare D. Weatherwise: Evaluation of a cue-based training approach for the recognition of deteriorating weather conditions during flight. Hum Factors. 2003; 45(2):337–345.
28. Wu S-C, Gooding C. L, Shelley A.E, Duong C.G, Johnson WW. Pilot convective weather decision making in en route airspace. Proceedings of the 28th International Congress of the Aeronautical Sciences (ICAS2012); 2012 Sept 23-28 Brisbane, Australia. Red Hook (NY): Curran Associates.