Training General Aviation Pilots for Convective Weather Situations

Elizabeth L. Blickensderfer; John M. Lanicci; Michael J. Vincent; Robert L. Thomas; MaryJo Smith; Jessica K. Cruit

BACKGROUND: Over the past 10-15 yr, considerable research has occurred for the development, testing, and fielding of real-time Datalink weather products for general aviation (GA) pilots to use before and during flight. As is the case with the implementation of most new technologies, work is needed to ensure that the users (in this case, the pilots) understand both the capabilities and limitations of the new technologies as well as how to use the new systems to improve their task performance. The purpose of this study was to replicate and extend a previous study on training pilots how and when to use these new weather technologies.

METHOD: This field study used a quasi-experimental design (pre- vs. post-test with a control group). There were 91 GA pilots from the Midwest, Northeastern, and Southeastern United States who participated in a 2-h short course or a control activity. The lecture-based short course covered radar basics, Next Generation Weather Radar (NEXRAD), NEXRAD specifics/limitations, thunderstorm basics, radar products, and decision making.

RESULTS: The pilots who participated in the course earned higher knowledge test scores, improved at applying the concepts in paper-based flight scenarios, had higher self-efficacy in post-training assessments as compared to pre-training assessments, and also performed better than did control subjects on post-test knowledge and skills assessments.

DISCUSSION: GA pilots lack knowledge about real-time Datalink weather technology. This study indicates that a relatively short training program was effective for fostering Datalink weather-related knowledge and skills in GA pilots.

KEYWORDS: aviation weather, NEXRAD, training.

Blickensderfer EL, Lanicci JM, Vincent MJ, Thomas RL, Smith M, Cruit JK. *Training general aviation pilots for convective weather situations*. Aerosp Med Hum Perform. 2015; 86(10):881–888.

It is no secret that weather can play havoc with general aviation safety.[8] Thunderstorms, in particular, are exceptionally dangerous, as they can produce high winds, shear, hail, icing, severe turbulence, and lightning. Despite improvements in aviation safety training and technology over the past several years, weather remains one accident factor that continues to contribute to fatalities in general aviation (GA).[2] While many sources of weather information are available to pilots, keeping up with rapidly changing aviation weather technology can be challenging. Requiring both an understanding of complex weather phenomena and knowing the capabilities and limitations of radars, correct interpretation of weather radar displays is not as straightforward as it appears to be on the evening news. Recent research offered a training course that helped young pilots to understand weather displays.[5] The purpose of the current study was to determine the efficacy of this training module using a sample that was more representational of the whole GA population in the United States.

GA pilots have multiple sources of information to use in identifying and avoiding storms. Web-based tools such as the Direct Access User Terminal System and the Aviation Digital Data Service are common preflight weather information sources, while the Telephone Information Briefing Service and Flight Service Stations provide preflight information over the phone. Weather information can also be accessed in flight, via radio broadcast, through the En-route Flight Advisory Service, the Hazardous In-flight Weather Advisory Service, or the Automated Surface/Weather Observing System stations.

From Embry-Riddle Aeronautical University, Daytona Beach, FL.

This manuscript was received for review in October 2014. It was accepted for publication in July 2015.

Address correspondence to: Dr. Beth Blickensderfer, Human Factors Department, Embry-Riddle Aeronautical University, 600 S. Clyde Morris Blvd., Daytona Beach, FL 32114-3900; blick488@erau.edu.

DOI: 10.3357/AMHP.4174.2015

Advancements in computer and communications technology over the last 10 yr have given GA pilots access to new in-cockpit weather products via real-time Datalink connections. Pilots are now able to view satellite imagery, pilot reports, aerodrome routine meteorological reports, Terminal Aerodrome Forecasts, and Next Generation Weather Radar (NEXRAD) feeds using panel-mounted displays or tablet computers paired with satellite or Automatic Dependent Surveillance - Broadcast receivers. All of these weather products have the ability to give pilots a detailed picture of the weather conditions surrounding their flight.

Despite the potential to improve the safety of flight, Datalink weather products—notably NEXRAD feeds—have limitations that can be deceptive to pilots who lack knowledge of these limitations. Data feeds can suffer from latencies as long as 20 min,[14] which can be hazardous if the pilot referencing the feeds is operating under the assumption that the radar is displaying precipitation in real-time. NEXRAD returns can also be distorted by distance from the radar site, atmospheric conditions, and geography.[14,19] Pilots can be lured into a false sense of security and find themselves in a dangerous situation without a route to safety.

Research regarding how pilots use cockpit weather information suggests that pilots may lack adequate understanding of both thunderstorm dangers as well as the limitations of Datalink weather products. These knowledge gaps could be fostering unsafe flight behaviors. For example, the Federal Aviation Administration's (FAA) guidance on the usage of Datalink weather displays stipulates that pilots use the displays for strategic weather decision making such as avoiding an area of severe weather completely, but not to navigate through an area containing severe weather cells, otherwise referred to as tactical decision making. Despite this guidance, a study performed by Latorella and Chamberlain[13] found that pilots believed that in-cockpit weather displays were appropriate for navigating through an area of severe weather. The study also found that the presence of an in-cockpit weather display reduced pilots' perceived risk associated with flying in areas of convective weather. Additionally, Beringer and Ball[4] indicated that pilots who viewed higher-resolution radar returns in a flight simulator were more likely to fly closer to severe cells than those who viewed low-resolution returns. Furthermore, pilots in this study who had a NEXRAD display spent significantly less time looking out the window and, in turn, could be at risk for missing important environmental cues.

Fortunately, research has indicated that weather-related training can positively impact pilots' use of Datalink weather displays. For instance, Wiggins and O'Hare[20] developed "Weatherwise," a cue-based training program that improved the timeliness of pilots' weather-related decisions. In another study, after receiving guidance on the correct usage of in-cockpit weather displays, pilots changed from tactical to a more strategic use of the displays.[3]

To address the Datalink weather knowledge gap further, Cobbett et al.[5] developed a training module to foster pilots' knowledge and skills for using NEXRAD-based Datalink weather products effectively in convective weather situations. Initial validation of the course demonstrated that the pilots who took the course gained knowledge and improved their performance on paper-based flight scenarios. While the study results were promising, the participating pilots' demographics [largely college students with low flight hours and associated with a Federal Aviation Regulation (FAR) Part 141/142 facility] were considerably different from typical GA pilots in the United States (middle aged, higher flight hours, and often trained under Part 61). Among other possible implications, the low flight hours may indicate that the pilots had encountered fewer flight situations involving convective weather, while the younger age may indicate a higher level of comfort with the new technology. Also, with all the pilots being at a FAR Part 141 facility, these pilots may have had more in-depth and recent flight and meteorological training than other GA pilots. Because of these differences, it is unclear whether the course would maintain effectiveness with more typical GA pilots.

The purpose of the current study was to examine the efficacy of the Cobbett et al.[5] training program to foster convective weather knowledge and skills of a more representative sample of GA pilots in the U.S. To accomplish this study, we sought GA pilots with three important differences from the pilots in the Cobbett et al. study. First, we aimed for pilots who were trained in FAR Part 61 facilities (rather than FAR Parts 141 or 142). Second, we wanted noncollegiate pilots (i.e., pilots who were not currently enrolled in a university). Finally, since the Cobbett et al. study was conducted in the Southeastern U.S. (an area with considerable convective weather activity), we also aimed to test pilots from a different geographic region. We focused on the Midwest and Northeastern U.S., as these regions have weather patterns different from the Southeastern U.S., but at the same time are subject to convective weather activity. Thus, pilots from these areas would have a higher probability of encountering convective weather while flying compared to pilots flying from the Pacific Northwest, for example.

METHODS

This field study used a quasi-experimental design (pre-test vs. post-test; and location/city). A supplemental analysis used a quasi-experimental mixed design (pre-test vs. post-test and control vs. experimental). The pre- vs. post-test comparison was to assess knowledge and skill gain following training as compared to before training. The addition of the control group was to rule out possible confounds, including history, testing, instrumentation, and maturation.[15]

All control group pilots were from the Central Florida data collection site. All other subjects were assigned to the experimental group. The reason behind this design was as follows. To participate, the GA pilots donated a whole day of their personal time to be in the study. Including control groups at each of the three locations would have meant that some subjects would have only performed a control activity and would not have received any training. This option did not seem to follow sound research ethics and the researchers decided to run a small control group in a location where it was feasible to also provide those subjects a follow-up session about using NEXRAD separate from the experiment. A lecture-based course, knowledge test, and paper-based scenarios were used during this experiment.

Subjects

Subjects were recruited from three cities. Recruiting methods were as follows: researchers posted flyers at local fixed base operators, contacted flying clubs and Civil Air Patrol squadrons and their members, and posted a listing on the FAA WINGS Program website. The study was approved in advance by the Embry-Riddle Aeronautical University Internal Review Board for the protection of human subjects, and each person provided written, informed consent prior to participation in the research.

In all, 94 pilots participated in the study; three subjects did not complete the post-test packet, leaving 91 pilots who completed the entire study. Of the subjects, 24 were from the Kansas City data collection site, 18 from Chicago, 32 from Boston, and 17 from Central Florida. There were 81 male subjects and 10 female subjects. The mean age was 54.6 yr (SD = 13.8) and ranged from 18 to 74 yr. The mean number of years flying was 19.6 (SD = 15.7) and ranged from 1 to 55 yr. The mean number of total flight hours was 1992 (SD = 3799; median = 520) and ranged from 50 h to over 25,000. The mean number of hours under instrument flight rules was 232 (SD = 467, median = 70).

Subjects were required to have obtained, at minimum, a private pilot certificate. There were 64 who possessed an instrument rating and 19 who were certificated flight instructors. Of the subjects, 66 received their private pilot certificate in a FAR Part 61 flight school, 11 subjects reported their private pilot training was performed in a FAR Part 141 flight school, and 14 subjects did not provide their flight training background. There were 30 subjects who reported they held a commercial certificate and 8 who reported holding an air transport pilot certificate. Approximately half of the subjects reported previously taking meteorology courses ($N = 42$) and/or using in-cockpit NEXRAD products ($N = 37$). Some subjects reported having taken meteorology courses in college ($N = 15$), attended aviation seminars ($N = 32$), or taken online weather courses ($N = 31$).

Equipment and Materials

The equipment for this study included the course given, a control activity, and evaluation instruments to assess the subjects' performance. The course used was developed by Cobbett et al.,[5] but the current study used an instructor who had not been part of the original course development. The course instructor was an FAA certificated airline transport pilot with airplane single engine land and airplane multi-engine land ratings. He also had a certification as a Gold Seal flight instructor with airplane single engine, airplane multi-engine, and instrument-airplane ratings and an advanced ground instructor.

The instructor gave a live, approximately 90-min PowerPoint® presentation. Topics covered in the lecture included radar basics, NEXRAD, NEXRAD specifics/limitations, thunderstorm basics, radar products, and decision making. At the end of each section the instructor posed questions related to the lecture to subjects to aid retention of course material.

As was done in the Cobbett et al.[5] study, paper-based practice scenarios were used to enable the subjects to practice applying the course content in an aeronautical decision making context. Scenarios began with a brief description of the purpose of a flight. Subjects were provided with preflight information, including time of departure, a map including all National Weather Service NEXRAD sites, and the applicable Terminal Aerodrome Forecasts, aerodrome routine meteorological reports, infrared satellite images, and single site, regional, and national NEXRAD mosaic images. The in-flight weather information was similar to what would be available with an in-cockpit weather product subscription service. The scenarios included questions that required subjects to apply the information given in the lecture to interpret the weather information correctly and make sound decisions regarding the flight.

The first practice scenario concerned a flight from Atlanta, GA, to Jacksonville, FL, to attend the Super Bowl, a flight that was scheduled to take 2 h. The weather along the planned flight path was clear with the exception of the Atlanta area, which was experiencing thunderstorms at the time of departure. Widespread thunderstorms formed along the planned route of flight 1 h into the flight. The instructor talked the subjects through this scenario, gave them time to answer the questions on their hard copies, and then gave feedback via a group discussion. The second practice scenario involved a 2.5-h flight from Sioux Falls, SD, to Burlington, IA, to attend a high school reunion. The preflight resources did not indicate any precipitation along the planned route of flight, but did portray a cold front positioned across the state of Iowa. After an hour into the flight, subjects had to decide whether or not to navigate through a line of thunderstorms that formed just north of the city of Burlington. The subjects worked through this scenario independently and, upon completion, the instructor led a group discussion about the scenario and provided feedback. The control activity was watching and discussing a video titled "Weather Wise."[1] The "Weather Wise" video included thought provoking considerations of aviation weather scenarios without specifically focusing on NEXRAD. After watching the video, the course instructor led a group discussion of the weather concepts in the video. Overall, the control activity was about convective weather, but the exact content and delivery was different from the NEXRAD course/module.

Evaluation Instruments

A demographics questionnaire gathered the subjects' ages, gender, state of residence, flight experience, and Datalink weather experience. The study also used multiple evaluation instruments: a radar knowledge test, scenario application tests, and attitudinal measures.

The radar knowledge test was a 28-item evaluation consisting of multiple-choice and true/false questions on the lecture concepts and terminology. The pre- and post-tests were nonidentical, parallel forms. Participant responses were recorded and overall percentage correct was calculated for the pre- and post-radar knowledge tests, respectively.

There were 2 paper-based 15-item scenario tests used to assess the participant's skills at interpreting the weather information in the context of a flight situation. The scenarios were of the same nature as those used in the practice portion of the course, but were different situations. The responses were scored for accuracy and a percent correct score was calculated for each participant.

The first scenario used in the pre- and post-tests involved a flight from Lake Charles, LA, to Montgomery, AL, to attend a funeral service for a family member. At the time of departure, the radar products showed a relatively clear flight path with thunderstorms in surrounding areas. The scenario stated that after 1 h into the flight, thunderstorms began to develop along the route of flight.

An additional, novel scenario was used in the post-test only. This second scenario was designed to be equivalent in complexity but with enough variation to appear different to the pilots. The second scenario included similar weather concepts and use of weather information displays, but the airports, location of the flights, and exact weather parameters differed. The purpose of examining subject performance on this second post-test scenario was to assess the degree to which the course fostered knowledge and skills that would generalize to situations other than only one aviation weather scenario. This scenario involved a flight from Lansing, MI, to Oshkosh, WI, to attend the Experimental Aircraft Associations' AirVenture fly-in and airshow. The route of flight required crossing Lake Michigan. The flight was clear of weather with the exception of the destination, where radar returns showed a large area of precipitation over the state of Wisconsin. After an hour into the flight, a squall line formed over Lake Michigan.

In addition to the knowledge and application tests, the study used two attitudinal measures: self-efficacy and trainee reactions. Higher self-efficacy has been identified as a positive outcome of training and self-efficacy has been shown to predict task performance.[6] To measure self-efficacy, the pilots completed a 10-item, 7-point Likert-type questionnaire (adapted from Riggs et al.[16]) that assessed the degree to which the subjects believed that they could use Datalink weather successfully. Cronbach's alpha was calculated to determine internal consistency for the pre- and post-tests (pre-test = 0.87, post-test = 0.83). Composite scores were then calculated by averaging the responses across the 10 items for each participant for both pre- and post-test self-efficacy questionnaires, respectively. Subjects rated the course utility and delivery on a 7-item, Likert style questionnaire (1 = low, 7 = high). Internal consistency was sufficient (Cronbach's alpha = 0.86) to justify computation of a composite score, and an average rating was computed for each subject.

Procedure

Subjects arrived at the respective testing sites and were provided with a consent form to review and sign. Once all of the subjects who had signed up to participate at a particular location had arrived, the research assistant briefed them about the research and instructions for completing the pretest packet. After the briefing was complete, the subjects were given 1 h to complete the pretest. If any of the subjects had not completed the pretest at the 1-h point, they were asked to stop. All subjects were then given a short break.

Next, for the experimental group, the instructor began the lecture portion of the training program. Once the lecture was completed, subjects practiced applying the concepts discussed during the lecture with the two practice scenarios. The procedure was the same for the control group with the exception that instead of the training lecture and scenarios, they watched the "Weather Wise" video,[1] discussed the video with the course instructor, and they did not perform the two practice scenarios. The total time spent in the course and practice scenarios was approximately 120 min, while the total time spent in the control activity was approximately 75 min.

Following course completion, the experimental group subjects had a 1-h lunch break. To help keep the time consistent with that of the experimental group, the control group subjects were given a 1.5-h lunch break following the control activity. After the lunch break, all subjects were given 1 h and 15 min to complete the post-test. Including the lunch break, the elapsed time between the pre-tests and the post-tests for the experimental group was approximately 180 min and for the control group it was approximately 165 min.

After subjects completed the post-test, the researcher debriefed each subject. Subjects then completed the research compensation form, were paid in cash, and were given the option to receive FAA WINGS credit.[9] Subjects were also given a "Radar Checklist" to use for future reference.

RESULTS

The first analysis focused on the subjects who received the training module ($N = 74$) and did not include the control group ($N = 17$) data. This was due to the following: first, the number of subjects in the control group was considerably less than those in the experimental group and the unequal sample sizes would have generated violations in assumptions for the statistical technique. Second, the control group was missing the attitudinal measures/data.

For the 74 subjects who received the training module, there were no significant differences between the three data collection locations in age, total number of flight hours, total number of instrument flight hours, total years flying, hours spent in weather radar training, or academic weather course credits earned. Hence the first analysis was a mixed-design multivariate analysis-of-variance (SPSS MANOVA, general linear model with repeated measures) with three dependent variables: radar knowledge, scenario knowledge, and self-efficacy. The between factor was location (Kansas City, Chicago, Boston) and the within factor was phase (pre-test vs. post-test).

Table I and **Table II** show the means and intercorrelations of the radar knowledge and scenario tests as well as the self-efficacy and reactions questionnaires, respectively. The inspection of the MANOVA using Wilk's Lambda showed that a significant main effect of the training occurred on the combined dependent measures [$F(3, 67) = 142.24$, $P \leq 0.001$, $\eta^2 = 0.87$]. It also indicated a significant effect of location [$F(6, 134) = 3.00$, $P = 0.009$, $\eta^2 = 0.12$]. There was not a significant interaction of training and location [$F(6, 134) = 0.76$, $P = 0.605$].

All subsequent univariate results reported are Greenhouse-Geisser statistics. First, the univariate tests revealed no significant effect of location on any of the dependent measures: the radar knowledge test [$F(2, 69) = 0.88$, $P = 0.42$]; the scenario knowledge [$F(2, 69) = 2.06$, $P = 0.14$]; and self-efficacy [$F(2, 69) = 0.24$, $P = 0.79$]. The likely reason that significance occurred in the omnibus test for location, but not in the follow-up univariate tests for location, was the unequal number of subjects at the different locations. In contrast, the univariate tests did reveal a significant effect of training on all three dependent measures and these will be discussed next.

First, results of the follow-up univariate tests showed a significant effect of training on radar knowledge [$F(1, 69) = 218.50$, $P \leq 0.01$, $\eta^2 = 0.76$]. As shown in Table I, subjects improved from an average pre-test score of 55% of questions correct to a post-test score of 77% correct. Furthermore, 76% of the variance in radar knowledge was accounted for by the training course. Next, the univariate test revealed a significant effect of training on scenario knowledge test scores [$F(1, 69) = 170.58$, $P \leq 0.01$, $\eta^2 = 0.71$]. Across all data collection sites, scenario test scores improved from 65% in the pre-test to 87% in the first post-test scenario, and 71% of the variability in the scenario-based test scores is accounted for by the training.

Since the second scenario was not included in the MANOVA due to lack of a separate pre-test, we examined the three scenario tests (scenario 1 pre-test, scenario 1 post-test, and scenario 2 post-test) with an additional one-way repeated measures ANOVA. The mean scores for all scenario tests are shown in Table I. The results indicated a significant difference between the three scores [$F(2, 72) = 94.32$, $P \leq 0.01$, $\eta^2 = 0.58$]. Post hoc examinations using a Bonferroni correction[18] revealed that the subjects scored significantly higher on the scenario 1 post-test than the scenario 1 pre-test [$t(73) = 13.04$, $P < 0.01$], higher on the scenario 2 post-test than the scenario 1 pre-test [$t(74) = 7.04$, $P < 0.01$], and higher on the the scenario 1 post-test than the scenario 2 post-test [$t(74) = 6.07$, $P < 0.01$].

The univariate test following the MANOVA revealed a significant effect of training on self-efficacy [$F(1, 69) = 94.32$, $P \leq 0.001$, $\eta^2 = 0.58$]. Responses to the self-efficacy questionnaire increased from an overall mean of 3.6 out of 7 on the pre-test to 4.8 out of 7 on the post-test. Of the variability in self-efficacy, 58% was accounted for by the training. This effect size is larger than some training studies in other areas.[10]

Regarding the learners' opinions about the course, the overall mean reaction score (out of 7) was 6.54 (SD = 0.511). This result indicates that the subjects had positive views of the course.

A separate set of analyses was conducted with the control group data. To generate equivalent sample sizes for this analysis, the researchers selected 17 subjects from the experimental group via a post hoc random selection process. First, the researchers removed any pilots with 10,000 or more flight hours from the existing experimental group. Then, using a random

Table I. Mean Test Score Before and After Datalink Weather Training Course.

	SCENARIO 1 PRE-TEST	SCENARIO 1 POST-TEST	SCENARIO 2 POST-TEST	
	M (SD)	M (SD)	M (SD)	*P*-VALUE*
Radar Knowledge (Percentage)				
Kansas City	56 (9)	79 (9)		
Chicago	55 (10)	73 (10)		
Boston	55 (14)	76 (14)		
Overall	55 (11)	77 (10)		≤0.01
Scenario Test (Percentage)				
Kansas City	66 (18)	87 (9)	80 (7)	
Chicago	67 (13)	92 (5)	78 (13)	
Boston	62 (16)	81 (18)	77 (16)	
Overall	65 (16)	87 (9)	79 (9)	≤0.01
Self-Efficacy				
Kansas City	3.36 (1.23)	4.82 (0.84)		
Chicago	3.71 (1.13)	4.79 (0.75)		
Boston	3.66 (1.2)	4.82 (0.95)		
Overall	3.59 (1.19)	4.80 (0.85)		≤0.01
Reactions				
Kansas City		6.41 (0.63)		
Chicago		6.75 (0.35)		
Boston		6.52 (0.46)		
Overall		6.54 (0.51)		

N = 74.

* *P*-values for follow-up univariate, pre vs. post comparison.

Table II. Correlation Matrix of Radar Knowledge, Scenario 1, Scenario 2, and Self-Efficacy.

VARIABLE	RADAR KNOWLEDGE POST-TEST *N* = 74	SCENARIO 1 POST-TEST *N* = 74	SCENARIO 2 POST-TEST *N* = 74	SELF-EFFICACY POST-TEST *N* = 73	SCENARIO 1 PRE-TEST *N* = 74
Radar Knowledge Post-test	-	0.71**	0.65**	0.38**	0.46**
Scenario 1 Post-test		-	0.72**	0.24*	0.43**
Scenario 2 Post-test			-	0.41**	0.21
Self-Efficacy Post-test				-	0.23*
Scenario 1 Pre-test					-

Pearson correlation coefficient: *correlation is significant at the 0.01 level (2-tailed); **correlation is significant at the 0.05 level (2-tailed).

number generator, the researchers selected 17 subjects from the original 74 subjects. These 17 subjects became the experimental group for comparison with the control group. To ensure equivalent groups, the researchers inspected the demographic data of these two groups and found no significant differences in terms of age, flight hours, instrument flight hours, years flying, rating, or FAR Part 61/141 training. To accommodate the relatively low total N ($N = 34$), separate t-tests were conducted on the three assessments. The means are shown in **Table III**. As noted previously, the control group data was missing the attitudinal measures (self-efficacy and reactions).

The radar knowledge pre-test scores did not differ significantly between the experimental and the control groups $[t(32) = 1.13, P = 0.27]$. Further, no between groups differences appeared for the scenario 1 pre-test scores $[t(32) = 0.08, P = 0.94]$. In contrast, the subjects who received training scored significantly higher than did the control subjects on the radar knowledge post-test $[t(32) = 4.72, P \leq 0.01]$, the scenario 1 post-test $[t(32) = 6.8, P \leq 0.01]$, and on the scenario 2 post-test $[t(32) = 3.10, P \leq 0.01]$. Also, the subjects who received the training had significantly higher delta scores (post-test minus the pre-test) on both the radar knowledge test $[t(32) = 3.01, P \leq 0.01]$ and the scenario 1 test $[t(32) = 5.29, P \leq 0.01]$. Thus, the subjects who received training improved from pre to post at a significantly higher rate than did the control group.

DISCUSSION

Convective weather continues to be a problem for GA safety. While real-time Datalink weather tools provide sophisticated meteorological information to pilots both prior to and during flight, if pilots do not use these tools effectively, it may actually cause more harm than good.[4] Training tools and strategies are needed to bridge the gap between Datalink weather technology and pilot performance using those tools.[11] The current field study provides evidence that a straightforward training course increased understanding of weather radar concepts and capability for application in convective weather scenarios in a highly generalizable sample of GA pilots. These results extend and replicate the previous effort to validate the training module.[5]

For effective training validation, it is key for training developers to consider learner or "training audience" characteristics, effects of the training on multiple measures of learning, and the transportability of the course to new instructors.[15] Examples of learner characteristics include age, prior experience, familiarity with the technology involved, educational background, and so on.[17] The current study used GA pilots from the Midwest and Northeastern United States with an average age in the mid-50s and most of whom had been trained under FAR Part 61 flight schools. This sample was quite different from that observed in the Cobbett et al.[5] study, and this sample is more typical of the GA pilot population in the United States. The fact that these pilots learned from the course indicated that other GA pilots would most likely learn from the course as well. Furthermore, the current findings indicate that the results of the Cobbett et al. study[5] were not based on characteristics of that sample (i.e., young, collegiate pilots). Additionally, training effectiveness researchers advocate that training validation use multiple measures of learning,[12,13,17] and the current study demonstrated strong training effects on several measures: knowledge of concepts, application of concepts, and attitudinal measures. Another necessity to ensure viability of a training program for

Table III. Summary of Study Results.

MEASURE	PRE-TEST (PERCENTAGE) *P* (*t*-TEST)	POST-TEST (PERCENTAGE) *P* (*t*-TEST)	DELTA SCORE M (SD) *P* (*t*-TEST)
Radar Knowledge	Exp = 57.5 Control = 52.9 $P > 0.05$	Exp = 75.8 Control = 59.2 $P \leq 0.01$	Exp. = 18.3 (11.0) Control = 6.3 (11.0) $P \leq 0.01$
Scenario I	Exp.= 65.3 Control = 65.6 $P > 0.05$	Exp. = 89.8 Control = 70.3 $P \leq 0.01$	Exp. = 24.5 (12.9) Control = 4.6 (8.5) $P \leq 0.01$
Scenario 2		Exp. = 78.2 (13.2) Control = 66.1 (9.2) $P \leq 0.01$	

$N = 17$ in each group.

widespread use is to determine that more than one instructor can give the course successfully. The current study demonstrated that a flight instructor who had not been involved in the original course design and development could teach the course effectively.

Field studies offer the opportunity to collect data from authentic users in their actual environments, but in doing so, these studies can lose some experimental control. The experimental design included several safeguards to protect against confounds: the control group, multiple data collection locations, and the second, novel post-test scenario. As described earlier, for ethical reasons, the researchers decided to run a small control group in a location where it was feasible to also provide those subjects with a follow-up instruction session on using NEXRAD (separate from the experiment). Even with the small control group, the current study results follow the same pattern as those of the Cobbett et al. study[5]—a large training effect in comparison to the control subjects. Furthermore, during the current study, researchers conducted the course in three different locations on three different dates and the pattern of results did not differ between sites—another indication that it was the module itself rather than another intervening variable which caused the post-test scores. Thirdly, subjects performed well on the novel, post-test scenario. Interestingly, they did not perform as well on the second scenario as they did on the first, which could mean that the pre-test itself had an effect on the post-test score. However, that subjects still performed well on the novel post-test scenario indicates that they had learned from the module and were able to generalize their knowledge and skills to a novel situation. Taken together, these results add to the likelihood that the module did indeed foster the subjects' learning gains as opposed to possible confounds such as a placebo effect, the course instructor, the pre-test, and subject maturation fostering the learning gains.

Similar to the Cobbett et al.[5] study, readers may question why the average scores were not higher on the assessments. Although large learning gains did occur, it may be possible to improve the course. With practice being essential to learning,[7] providing additional practice scenarios could help. Other reasons underlying the moderate post-test scores could be measurement/test construction issues such as poor questions (e.g., difficult for the respondents to interpret correctly even if they have a high understanding of the concepts), highly difficult questions (e.g., questions requiring respondents to understand detailed nuances of the concepts), and/or items that were low in content validity (i.e., did not cover adequately the content presented in the course either by testing concepts that were not discussed in the course and/or did not adequately address concepts that were included in the course). Further examination via a statistical item analysis of the assessment instruments used in this study indicated that measurement issues could be responsible for post-test scores reaching only a moderate level.

Paralleling other research,[4,15,20] the pre-test scores in the current study indicate that GA pilots have knowledge gaps regarding concepts underlying the effective use of NEXRAD-based products in convective weather situations. Fortunately, experiencing the course described in this study helped pilots respond correctly to convective weather questions and scenarios. Future research should include pilots performing simulated flights. In this way, we could determine the impact of the course on actual flight performance. In the meantime, however, combining the results of the current study with the Cobbett et al.[5] results gives ample evidence that this course fosters knowledge of using NEXRAD-based products. Thus, it is our hope that as many interested pilots as possible have the opportunity to take this course and subsequently use their increased knowledge and skill of NEXRAD-based tools to avoid being caught in unsafe situations.

ACKNOWLEDGMENTS

This research was conducted under sponsorship of the Federal Aviation Administration's (FAA) Center for Excellence in General Aviation Research under FAA Cooperative Agreement Number 07-C-GA-ERAU-014.

A portion of this data was presented at the annual meeting of the Human Factors and Ergonomics Society 2013 in San Diego, CA.

The views expressed in this paper are those of the authors and do not represent those of the organization with which they are affiliated or the FAA. The authors wish to thank the hosts of our data collection sites and all of the study subjects.

Authors and affiliations: Elizabeth L. Blickensderfer, M.S., Ph.D., Human Factors Department, John M. Lanicci, M.S., Ph.D., Department of Graduate Studies, Robert L. Thomas, B.S., M.S., Aeronautical Science Department, MaryJo O'Connell Smith, M.Ed., Ph.D., and Jessica K. Cruit, M.A., M.S., Embry-Riddle Aeronautical University, Daytona Beach, FL, and Michael J. Vincent, B.A., M.S., NASA Langley Research Center, Hampton, VA.

REFERENCES

1. Aircraft Owners and Pilots Association. 2007. Weather wise: practical tips and tactical tricks [DVD]. [Accessed 7 July 2015]. Available from http://www.asf.org.
2. Aircraft Owners and Pilots Association. 2010. 2010 Nall report: accident trends and factors for 2009. AOPA Air Safety Foundation. [Accessed March 12, 2015]. Available from http://www.aopa.org/-/media/Files/AOPA/Home/News/All-News/2012/April/VFR-in-to-IMC-Learn-to-escape-the-odds/10nall.pdf.
3. Ball JD. 2008. The impact of training on general aviation pilots' ability to make strategic weather-related decisions. [Accessed March 12, 2015]. Available from http://www.dtic.mil/cgibin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA477162. FAA Report no. DOT/FAA/AM-08/3.
4. Beringer DB, Ball JD. 2004. The effects of NEXRAD graphical data resolution and direct weather viewing on pilots' judgments of weather severity and their willingness to continue a flight. [Accessed March 12, 2015]. Available from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA423239 &Location=U2&doc=GetTRDoc.pdf. FAA Report no. DOT/FAA/AM-04/5.
5. Cobbett EA, Blickensderfer E, Lanicci JM. Evaluating an education/training module to foster knowledge of cockpit weather technology. Aviat Space Environ Med. 2014; 85(10):1019–1025.
6. Davis WD, Fedor DB, Parsons CK, Herold DM. The development of self-efficacy during aviation training. J Organ Behav. 2000; 21(8):857–871.

7. Ericsson KA, Prietula MJ, Cokely ET. The making of an expert. Harv Bus Rev. 2007; 85(7-8):114-121.
8. Federal Aviation Administration, Aviation Safety Information Analysis and Sharing. 2010. Weather-related aviation accident study 2003-2007. [Accessed March 12, 2015]. Available from http://www.asias.faa.gov/pls/apex/f?p=100:8:0::NO::P8_STDY_VAR:2.
9. Federal Aviation Administration. 2011. WINGS: pilot proficiency program. Advisory Circular AC 61-91J. [Accessed March 12, 2015]. Available from http://www.faa.gov/documentLibrary/media/Advisory_Circular/AC%2061-91J.pdf.
10. Gist ME, Schwoerer C, Rosen B. Effects of alternative training methods on self-efficacy and performance in computer software training. J Appl Psychol. 1989; 74(6):884–891.
11. Kirk LF, Martin L, Lanicci JM, Frushour GV, Derry J, et al. A general aviation perspective for the Weather Technology In the Cockpit Program. Washington (DC): DOT/FAA/Aviation Weather Group Air Traffic Organization; 2011.
12. Kraiger K, Ford JK, Salas E. Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. J Appl Psychol. 1993; 78(2):311-328.
13. Latorella KA, Chamberlain JP. Tactical vs. strategic behavior: general aviation piloting in convective weather scenarios. Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting. Sept. 30-Oct. 4, 2002; Baltimore, MD. Santa Monica (CA): HFES; 2002:101-105.
14. National Weather Service Modernization Committee, Commission on the Engineering and Technical Systems, National Research Council. 1995. Assessment of NEXRAD coverage and associated weather services. [Accessed July 23, 2015]. Available from: http://www.nap.edu/openbook.php?record_id=9056.
15. Noe RA. Employee training and development, 6th ed. Boston (MA): McGraw-Hill/Irwin; 2012.
16. Riggs ML, Warka J, Babasa B, Betancourt R, Hooker S. Development and validation of self-efficacy and outcome expectancy scales for job-related applications. Educ Psychol Meas. 1994; 54(3):793–802.
17. Salas E, Tannenbaum SI, Kraiger K, Smith-Jentsch KA. The science of training and development in organizations: What matters in practice. Psychol Sci Public Interest. 2012; 13(2):74–101.
18. Toothaker LE. Multiple comparison procedures. Newbury Park (CA): Sage; 1993.
19. University of Illinois. 1999. Department of Atmospheric Sciences, World Weather 2010 Project, Online Guides. [Accessed July 23, 2015]. Available from http://ww2010.atmos.uiuc.edu/(Gh)/guides/home.rxml.
20. Wiggins MW, O'Hare D. Expert and novice pilot perceptions of static in-flight images of weather. Int J Aviat Psychol. 2003; 13(2):173–187.