5-7-2019

A Tablet-Computer App Displaying Runway Winds

William R. Knecht

A TABLET-COMPUTER APP DISPLAYING RUNWAY WINDS

William R. Knecht
Civil Aerospace Medical Institute, FAA,Oklahoma City, OK, USA

We tested variants of a mobile meteorological tablet-computer application designed to help general aviation (GA) pilots land aircraft more safely under windy conditions. This "app" compared METAR runway wind information in several graphical and textual formats. Study 1 tested 25 GA pilots on 18 runway wind scenarios. Graphical METARs depicted the runway with a large arrow at 90°, representing the crosswind speed component, and a second arrow parallel to the runway, representing the headwind/tailwind component. We hypothesized that eliminating the need for complex mental calculation of wind components would increase speed and/or accuracy of information processing. Study 2 tested 17 pilots on 24 scenarios, employing the same basic method, but enhanced by color-coding the wind-component arrows according to each pilot's previously stated maximums for landing wind risk-tolerance. Both studies showed that runway-relative, two-arrow wind component depictions were significantly fastest and most efficient. Pilots unanimously preferred graphical displays over textual.

Adverse winds are a persistent challenge for all pilots and, therefore, a high priority for the FAA (FAA, 2017). Winds at landing are particularly problematic to general aviation (GA).

The current research continues empirical testing of a low-cost, portable GA device designed to deliver timely weather information to the flight deck. This mobile meteorological application runs on a tablet computer (iPad), and is currently under development by the Research Applications Laboratory (RAL) of the National Center for Atmospheric Research (NCAR, Ahlstrom, Caddigan, Schulz, Ohneiser, Bastholm, & Dworsky, 2015, Knecht & Dumont, 2019, Knecht & McCarthy, 2019).

Common Method of Experiments 1 and 2

Measuring "Quality" of Information Depiction

This "app" can present runway wind information similar to that shown in Figure 1. The research question centered on finding the best type of information to display for that purpose.

In order to support a claim that depicting wind information one way is "better" than another there has to be some method of objectively quantifying display *quality*. The metrics of quality measured here were *accuracy* and *speed* of the pilots' mental wind-evaluation process.

Decision *speed* was simply how much time it took the pilot to decide whether or not to land, given the runway-level wind information shown. Decision *accuracy*, however, was an altogether-different and harder quality to assess. To assess accuracy, we compared "objective landing difficulty" to "perceived landing difficulty" on the assumption that *the closer the* ***perceived difficulty*** *of a wind scenario was to its* ***objective difficulty****, the better the wind display*. Figure 2 explains.

Creating Scenarios with Known Objective Difficulty

Operationalizing the experimental method required wind scenarios with various objective levels of difficulty. This required controlling for each pilots' *skill* and *risk-tolerance*. For instance, if one pilot thought a 3-kt crosswind was "easy" and another thought a 5-kt crosswind

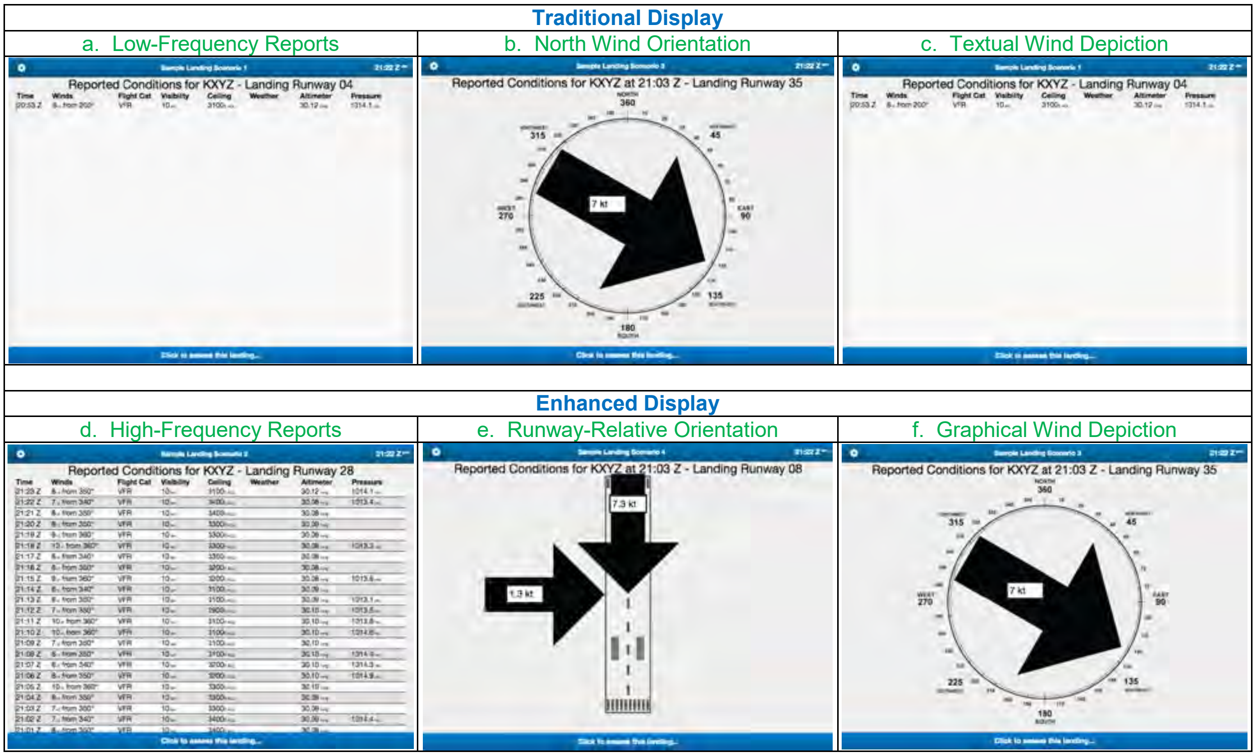


Figure 1. NCAR's Experiment 1 six runway wind depictions, all samples being supposedly 19 minutes old: a) "Traditional," text-based, similar to an aviation routine weather report (METAR); b) "Traditional" graphical wind depiction, a north-up view with an arrow showing wind direction and textual depiction of speed; c) "Traditional" textual METAR; d) "Enhanced" information similar to "a" but updated each minute, with the newest information no more than 1 minute old; e) "Enhanced" graphical wind depiction, a runway-relative view with separate arrows for crosswind and runway-aligned wind components, and; f) "Enhanced" METAR, similar to "c," but graphical as "b."

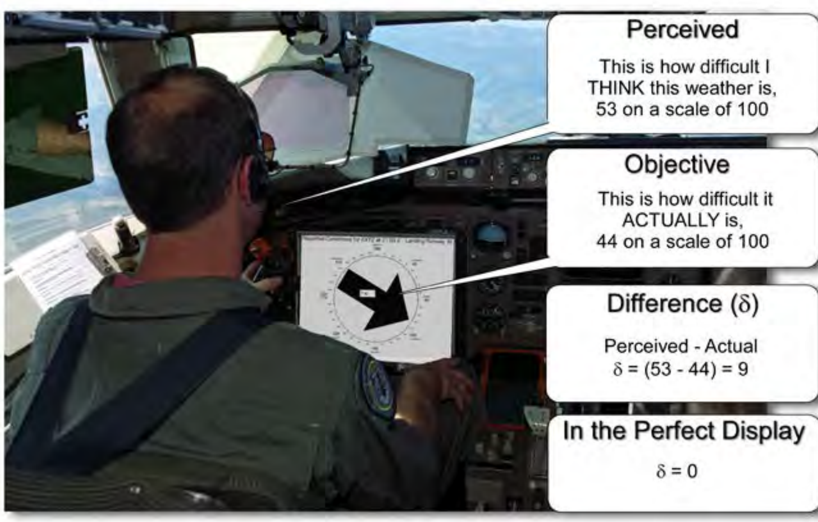


Figure 2. "Display quality" was measured as a difference score δ (delta), defined as participant's perceived scenario difficulty minus their objective scenario difficulty, both on a scale of 0—100. In a "perfect" display δ would equal zero; the display enabled them to correctly assess the scenario difficulty. Long, dry runway was assumed here.

was easy, to construct an objectively "easy" scenario, we would obviously want the crosswind component to be between 0 and 3 kts for the first pilot and 0—5 kts for the second pilot. This kind of individual adjustment is called *normalization*, and its goal is to create a single "normal" scale (e.g., 0—100 "difficulty units") that can be applied to all pilots, no matter what their skill or risk tolerance. This then allows statistical comparisons across experimental conditions.

To create such a "normal scale," at the very beginning of each pilot's test session we had pilots give us their individual "thresholds" for wind-component speeds. "Low Threshold" was defined as "Below that speed I wouldn't worry about that wind component." "High Threshold" was defined as "Above that speed I would hesitate to land with that wind component." Then, knowing each pilot's "easy" and "difficult" wind speeds, we could objectively define "easy" and "difficult" scenarios for each individual pilot. Additionally, from these two values we could interpolate a "moderate" difficulty by simply picking a value halfway between the two extremes.

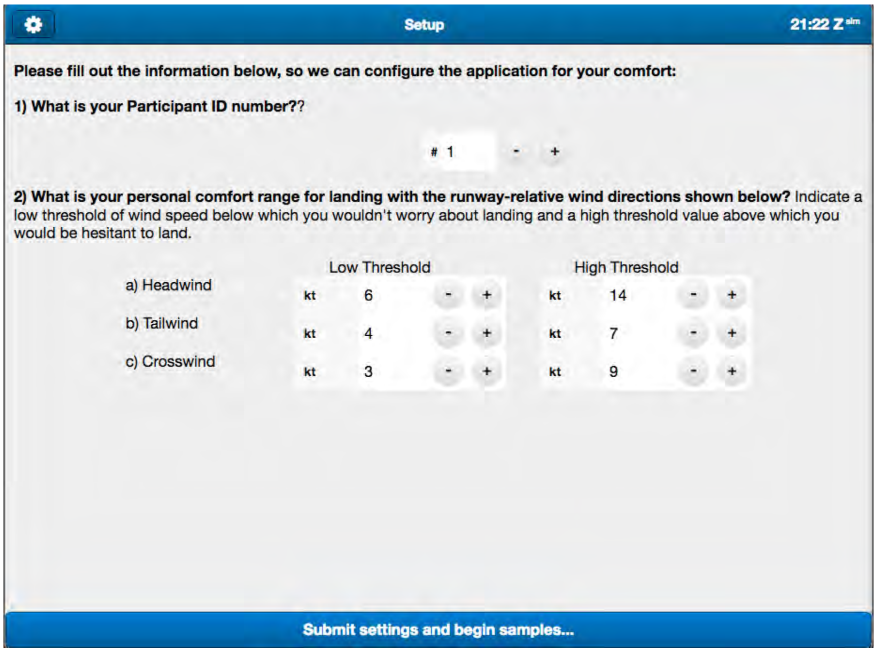


Figure 3. 1- Screenshot of the Setup page, showing the example of a "Low Headwind Threshold" of 6 kt and a "High Crosswind Threshold of 9 kt" for a hypothetical pilot.

Assessment of "Decision Quality"

Measuring *Decision Speed* was straightforward. This was merely the time it took each pilot to assess the wind situation, defined as the elapsed time from when the wind information page was first shown until the instant the pilot moved on to the subsequent assessment page.

To measure *Decision Accuracy*, pilots were asked to indicate each scenario's *perceived landing difficulty* by moving sliders along the "normal scale" of 0-100 (Fig. 4), representing how difficult the landing was *expected* to be. Meanwhile, recall that each scenario's *objective landing difficulty* had been normalized for that pilot, based on her/his previously reported values for how wind speed and direction would affect landing difficulty for him or her, personally. Therefore, the assessment page gave everything else necessary to calculate *perceived – objective difficulty* = δ. And, if one wind depiction was truly higher-quality than another, we would expect most of the δ scores to be smaller.

Experiment 1

Experimental Design

Experiment 1 utilized a within-participants (repeated measures) statistical design. Each pilot responded to 18 runway wind landing scenarios, each depicted by a single page similar to

Figure 1's, with a different set of wind parameters as independent variables. Figure 5 illustrates.

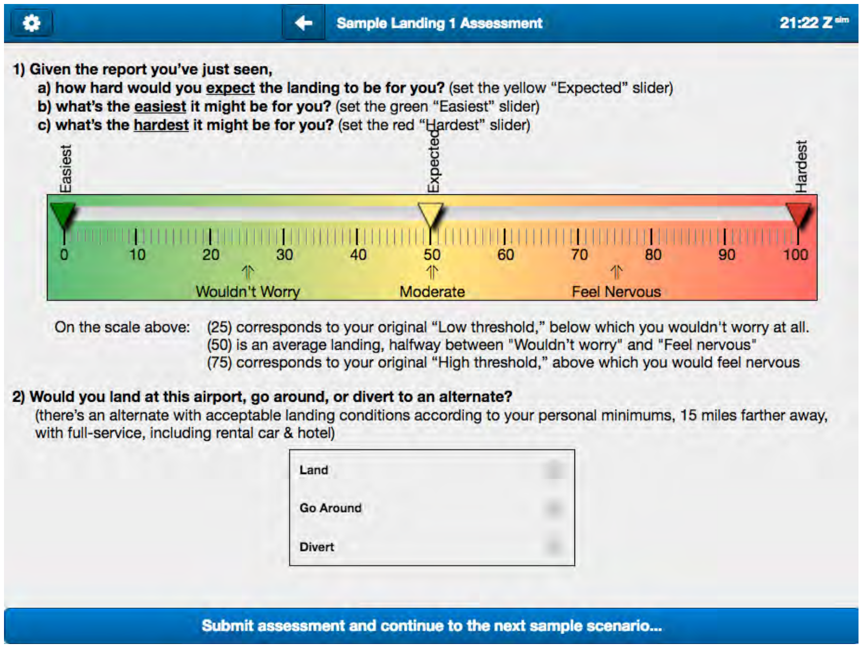


Figure 4. Screenshot of the Evaluation page.

Research design, **2 x 3 x 3** (Display Type (A) x Information Type (B) x Scenario Difficulty (C))						
	A1-Enhanced Display			**A2-Traditional Display**		
Scenario Difficulty[A]	B1-High-Frequency Reports	B2-Runway-Relative Orien-tation	B3-Graphical Wind Depiction	B1-Low-Frequency Reports	B2-North-Up Orientation	B3-Textual Wind Depiction
C1-Easy						
C2-Moderate						
C3-Difficult						
[A]All scenarios' objective difficulties were set according to individual pilots' answers for "Low Threshold" and "High Threshold" values on their Setup page (see text for details).						

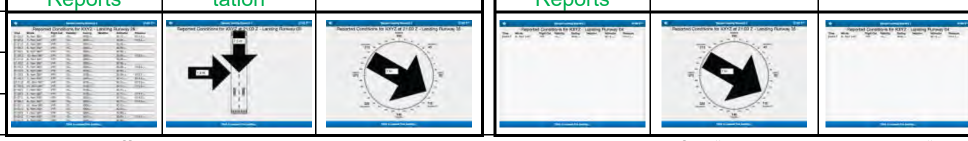


Figure 5. Experiment 1's 2×3×3 research design.

Participants

Twenty-four GA pilots were recruited from a local flight school and paid $50 (Fig. 6).

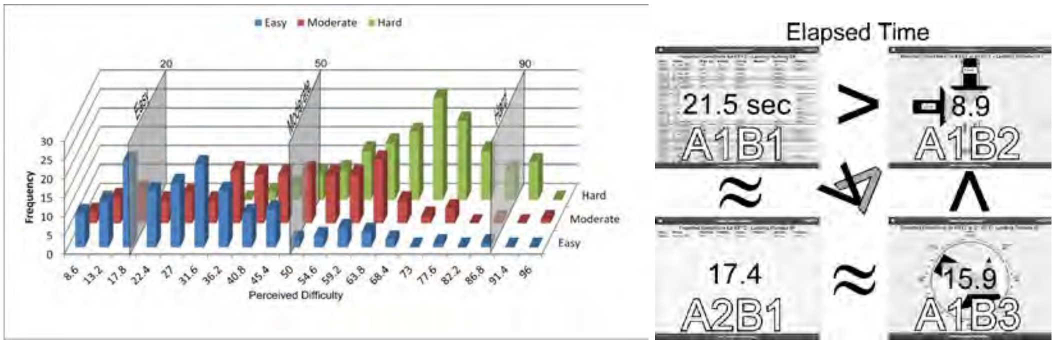
Student	0	**CFII**	5	**Age-mean**	28.8	**TFH[1]-mean**	959
Private pilot	24	**Commercial**	8	**Age-median**	21.5	**TFH-median**	185
Instrument-rated	9	**ATP**	2	**Age-SD**	14.6	**TFH-SD**	2150
CFI	7	**Multi-engine**	7	[1] Total Flight Hours			

Figure 6. Experiment 1 pilot demographics.

Results

Overall 2×3×3 ANOVA analysis of *Perceived Scenario Difficulty* δ scores showed significance only for the three objective difficulty levels (C1-3). Pairwise post-hoc comparisons indicated that each of those three levels was perceived significantly different from the other two at $p = .00001$ or better. However, as Figure 7a shows, there was considerable spread in the data.

The graphical twin-arrow display, depicting separate crosswind and headwind/ tailwind components, was fastest, with no apparent loss of accuracy representing landing difficulty.



a b

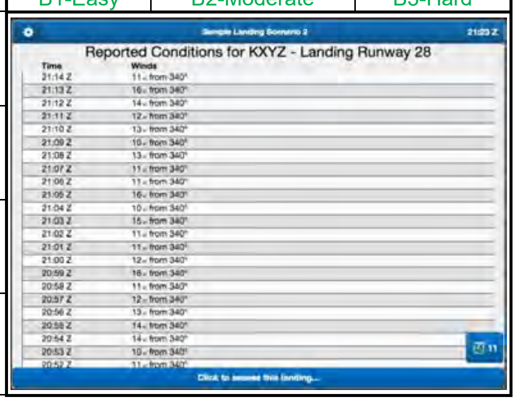
Figure 7. Experiment 1. a) variation in *Perceived Landing Difficulty* across Easy, Moderate, and Hard Scenarios, b) pairwise landing decision speeds (note that twin-arrow (A1B2) was fastest.

Experiment 2

Experimental Design

Experiment 2 leveraged the results of Experiment 1. Figure 8 illustrates. In a 2×3×4 repeated measures design, the number of depictions (A1-2) was reduced to two and the two-arrow depiction was color-coded to represent objective landing difficulty. Red represented a wind component speed greater than pilot's pre-stated maximum tolerance, orange represented medium-concern speeds and green represented "no worry."

Research design, 2 x 3 x 4 (Display Type (A) x Scenario Difficulty (B) x Time Constraint (C))						
	A1-Textual Display			A2-Graphical Display		
Time Constraint	B1-Easy	B2-Moderate	B3-Hard	B1-Easy	B2-Moderate	B3-Hard
C1-40 seconds						
C2-20						
C3-10						
C4-5						



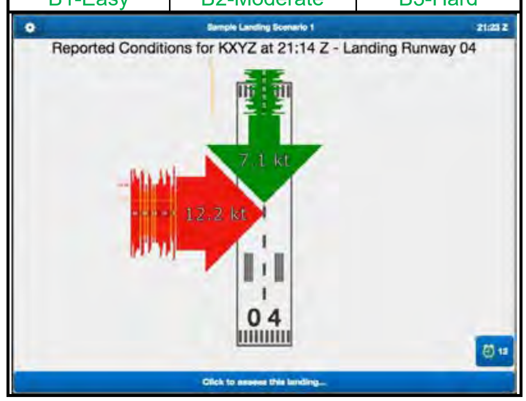


Figure 8. Experiment 2's 2×3×4 research design.

Again, three levels of scenario difficulty (B1-3) were used. Four levels of time constraint (C1-4) were introduced to see how restricting available viewing time would affect performance.

Participants

Seventeen GA pilots were recruited from a local flight school and paid $50 (Fig. 9).

Student	0	**CFII**	3	**Age-mean**	22.3	**TFH-mean**	323
Private pilot	17	**Commercial**	7	**Age-median**	22.0	**TFH-median**	200
Instrument-rated	15	**ATP**	0	**Age-SD**	3.4	**TFH-SD**	205
CFI	4	**Multi-engine**	4	**TFH-max**	800	**TFH-min**	98

Figure 9. Experiment 2 pilot demographics.

Results

Introduction of *Time Constraint* resulted in severe data non-normalities in *Perceived*

Landing Difficulty, disallowing ANOVA. Figure 10 shows *p*-values and effect sizes for paired t-tests, normality permitting, with Wilcoxon *p*-values for variable pairs involving a non-normality.

Significances between IV pairs (and Cohen's *d* effect size)																
	A2	B2	B3	C2	C3	C4		A1	A2	B1	B2	B3	C1	C2	C3	C4
A1	.337 (.69)								$7.21*10^{-5}$ (2.92)							
B1		$1.028*10^{-10}$ (15.10)	$1.452*10^{-11}$ (16.97)								.004 (1.59)	.149 (.69)				
B2			0.025 (2.34)									.088 (.93)				
C1				.412 (.47)	.660 (.25)	.590 (.36)								.040 (.99)	$2.88*10^{-4}$ (3.44)	$2.93*10^{-4}$ (5.32)
C2					.742 (.23)	.853 (.09)									$1.55*10^{-4}$ (3.14)	$2.93*10^{-4}$ (5.89)
C3						.818 (.12)										$5.03*10^{-4}$ (4.89)
								9931	6942	7571	9401	8338	11888	10251	6977	4632
								IV group means (milliseconds)								

Figure 10. Experiment 2 *p*-values and effect sizes for *Perceived Landing Difficulty* and *Elapsed Viewing Time*.

Elapsed Viewing Times were significantly different for graphical depictions (A1 vs A2). Differences between levels of *Time Constraint* were significant but logically trivial. More meaningful was that, given 40 seconds (C1), only one pilot timed-out on one scenario, whereas 74% of scenarios timed-out when pilots had only 5 seconds (C4).

Conclusions

These two studies clearly showed that, even when time is short, pilots *can* often discriminate between difficult and easy runway winds using either textual or graphical wind displays. However, this seems to be because they use a shortcut, or *heuristic*, when pressed for time. Rather than mentally computing wind components, they simply scan for wind speeds higher than their comfort level, regardless of wind direction. This allows quick scan of even long columns of numbers. But, deriving wind *components*—particularly intermediate-difficulty components—is a far more difficult task, particularly when time is short.

We therefore suggest that medium- and high-difficulty wind components will be best depicted by a graphical two-arrow display, particularly one color-coded according to each pilot's personal maximums reflecting their skill and risk-tolerance within the context of a given aircraft. Pilots here concurred, unanimously preferring the graphical displays over the textual.

Acknowledgments

This research was funded by the FAA's Weather Technology in the Cockpit Program, ANG-C61. Arnaud Dumont and Paddy McCarthy of NCAR's Research Applications Laboratory helped design the study and coded the mobile application.

References

Ahlstrom, U, Caddigan, E., Schulz, K., Ohneiser, O., Bastholm, R., & Dworsky, M. (2015). *Initial assessment of portable weather presentations for general aviation (GA) pilots*. (Technical Report DOT/FAA/TC-15/42).

Knecht, W.R., & Dumont, A. (2019). *Tailoring surface winds information for mobile meteorological applications: Part I, beta-testing*. (OAM technical report, in review).

Knecht, W.R., & McCarthy, P. (2019). *Mobile meteorological information tailored to landing phase of flight. Part II: Refinement*. (OAM technical report, in review).