The Effectiveness of Airline Pilot Training for Abnormal Events

Stephen M. Casner, National Aeronautics and Space Administration, Moffett Field, California, **Richard W. Geven**, and **Kent T. Williams**, San José State University Research Foundation, San Jose, California

Objective: To evaluate the effectiveness of airline pilot training for abnormal in-flight events.

Background: Numerous accident reports describe situations in which pilots responded to abnormal events in ways that were different from what they had practiced many times before. One explanation for these missteps is that training and testing for these skills have become a highly predictable routine for pilots who arrive to the training environment well aware of what to expect. Under these circumstances, pilots get plentiful practice in responding to abnormal events but may get little practice in recognizing them and deciding which responses to offer.

Method: We presented 18 airline pilots with three abnormal events that are required during periodic training and testing. Pilots were presented with each event under the familiar circumstances used during training and also under less predictable circumstances as they might occur during flight.

Results: When presented in the routine ways seen during training, pilots gave appropriate responses and showed little variability. However, when the abnormal events were presented unexpectedly, pilots' responses were less appropriate and showed great variability from pilot to pilot.

Conclusion: The results suggest that the training and testing practices used in airline training may result in rote-memorized skills that are specific to the training situation and that offer modest generalizability to other situations. We recommend a more complete treatment of abnormal events that allows pilots to practice recognizing the event and choosing and recalling the appropriate response.

Application: The results will aid the improvement of existing airline training practices.

Keywords: abnormal events, generalizability, effectiveness, rote learning, training, pilots

Address correspondence to Stephen Casner, NASA Ames Research Center, Mail Stop 262-4, Moffett Field, CA 94035; e-mail: stephen.casner@nasa.gov.

A number of aircraft accident reports describe situations in which pilots encountered an abnormal event for which they had rehearsed many times before. But instead of offering the well-practiced response learned during training, pilots did something different (NTSB, 1995, 2010a, 2010b). Given that each pilot had demonstrated their ability to correctly respond to these events on numerous occasions, the accidents leave us with the question of what went wrong. With almost 10,000,000 U.S. airline departures conducted each year by more than 80,000 pilots, we might wonder if these events are simply rare and unfortunate exceptions. On the other hand, to learn that difficulties with the real-world application of these skills were more widespread would naturally draw our attention to the way that we train pilots to respond to these unusual events.

Most jurisdictions require that air carrier pilots periodically take time off from their everyday flying jobs to complete a simulator training program that focuses almost entirely on abnormal events that rarely happen but for which it is important to remain prepared. During these few days, pilots are presented with rarities such as aerodynamic stalls, encounters with hazardous weather, and engine failures. The lineup of events that pilots see during these few days is rational: Regulators choose those events that present the greatest chance of occurrence, that present the greatest challenges to pilots, and that impose the greatest penalties for getting them wrong. Regulations additionally set forth specific test standards that describe an acceptable performance on the part of the pilot. On first glance, this plan to ensure pilot readiness for the unexpected seems to be logically designed.

Presented with a list of abnormal events, airline companies in turn devise standard operating procedures (SOP) that are rigorously practiced by pilots when responding to these

events during simulator sessions. But since they are required by regulation to be practiced and tested, pilots know that each abnormal event will occur at some point during the training session. Since airline companies stop short of creating customized testing regimes for their many pilots, all pilots typically see the same sequence of abnormal events presented under the same circumstances. Furthermore, since their jobs depend on successfully completing this required training and testing, discussions of which events to expect at which times and under which circumstances during the training and testing sessions are, needless to say, popular among pilots. These circumstances combine to make airline training events highly scripted and predictable exercises that call into question the extent to which pilots' abilities to recognize and respond to abnormal events, in all the forms in which they might present themselves during a real flight, are being honed and tested.

Previous research has pointed out ways in which the training practices described previously could lead to the sort of difficulties in responding to abnormal events described by accident reports. First, many studies have demonstrated that skills that are taught and practiced in a rote manner (using the same presentation each time) are less likely to generalize or transfer than those taught in a more varied or meaningful way (Healy, Schneider, & Bourne, 2012; Kieras & Bovair, 1984; Meredith, 1927; Singley & Anderson, 1989; Taatgen, Huss, Dickison, & Anderson, 2008; Woodrow, 1927). These studies suggest that such practices can leave learners with narrow, memorized understandings of problem situations that do not generalize well to situations that do not match the ones they see in training. Second, because of their rarity, unusual events naturally contain an element of surprise: a psychological condition in which one's cognitive capabilities may be, in the throes of excitement, temporarily compromised (Hancock & Szalma, 2008). Presenting pilots with abnormal events in highly predictable ways (with the surprise removed) may again deny them opportunity to practice recognizing and responding to these events under less expected circumstances.

The first goal of our study was to determine if the difficulties in responding to abnormal

events described by a handful of accident reports were indicated in a more random sample of airline pilots. A second goal was to link any difficulties observed in responding to abnormal events to the methods used to teach and practice these events during airline training. Toward this end, we designed a simple experiment in which we presented pilots with abnormal events in two different ways: (a) under the predictable circumstances used during airline training (our control condition) and (b) under less predictable circumstances, as they might occur during a real flight (our treatment condition). If we observe little difference in pilot performance between these two conditions, we might feel confident that (acknowledging a few rare exceptions described in accident reports) existing training methods help produce robust skills that will serve pilots in a variety of situations. On the other hand, if we observe significant differences in performance, we might consider the idea that current training methods simply prepare pilots to pass a particular and very familiar test.

METHOD

Participants

For this study, 18 active Boeing 747-400 pilots, 9 captains and 9 first officers, participated on a voluntary basis. Pilots ranged between 5,000 and 20,000 hours of total flight experience ($M = 11,056$ hours, $SD = 3,670$). Pilots had accumulated an average of 356 hours during the past 12 months ($SD = 178$). Of the subjects, 6 pilots completed their initial training in the military while 12 pilots were trained in the civilian (general aviation) environment.

Apparatus

The Boeing 747-400 (Level D) flight simulator located at the NASA Ames Research Center was used for the experiment. The checklists, quick reference handbook (QRH), and avionics configuration used by each employer of each pilot were provided for each simulator session.

The Abnormal Events

All 18 pilots were presented with three abnormal events that are required by regulation to be practiced during airline training: (a) aerodynamic stall, (b) low-level wind shear, and (c)

engine failure on takeoff. Each of these three events was presented to pilots in two different ways: (a) under the circumstances used during airline training and (b) under unexpected circumstances. The following describes the details of the abnormal events that each pilot in our experiment saw.

Aerodynamic stalls. Pilots were presented with three stalls: one familiar stall practiced during training and two less expected stalls.

As a control condition, pilots were asked to demonstrate a power-off stall. This demonstration of a stall and stall recovery is the one practiced by pilots during every airline training and testing event. During the stall demonstration, the pilot is asked to retard the throttle levers and intentionally allow the speed of the airplane to decay. As the airplane approaches a stalled condition, the stick shaker is activated. As soon as the stick shaker activates, the pilot immediately moves the throttle levers forward to maximum thrust. A defining characteristic of the stall demonstration is that not only do pilots know that the stall is imminent, the stall is incurred as a result of their own deliberate actions. The stall demonstration is always practiced at low altitudes (typically 10,000 feet).

As a first variation on the stall demonstration, pilots were presented with a stall as they reached an altitude of 2,500 feet while climbing out after a routine takeoff. This stall was created by the experimenters by rapidly shifting the prevailing winds from a strong headwind to a light tailwind condition. This wind change caused the airplane to quickly lose airspeed, experience a stall buffet and associated stick shaker, abruptly pitch down, and begin to rapidly descend. The correct response is the same response used during the stall demonstration detailed previously.

As a second treatment condition, pilots were presented with a stall as they descended through 34,500 feet during a routine descent. Again, the stall was created by the experimenters by rapidly shifting the prevailing winds. Again, the correct response to this situation includes immediately applying maximum thrust.

Low-level wind shear. Pilots were presented with two low-level wind shear encounters during approach to landing: one expected and one unexpected.

As a control condition, pilots were presented with a low-level wind shear event as it is practiced during airline training and testing. During these training and testing events, pilots are aware that a wind shear encounter must be given at some point during the simulation period. In the early portion of one of the arrivals that pilots perform during training, they will receive a weather observation that explicitly warns them about conditions conducive to wind shear and a report about a previous aircraft that experienced wind shear during their approach. From this point on, pilots are well aware that the required wind shear encounter is about to be practiced. Upon encountering the wind shear, an automated system designed to detect and alert pilots to wind shear sounds a warning in the cockpit ("*Wind Shear! Wind Shear!*"). The correct procedure for escaping a wind shear encounter is to abandon the approach, advance the throttle levers to maximum power, pitch the airplane to the target pitch displayed on each pilot's instruments, and refrain from retracting the landing gear or flaps until the airplane has completely escaped the wind shear zone. In our experiment, pilots encountered the wind shear as they descended through 600 feet (above ground level) and were about to land.

As a treatment condition, pilots were presented with the same low-level wind shear encounter without the advance warnings. As pilots began their approach they received a weather observation that gave no indication of wind shear conditions. Upon descending through 600 feet, pilots encountered the wind shear. During this event, the experimenters had disabled the auditory alerting system and pilots were left to recognize the wind shear encounter on their own.

Engine failure on takeoff. Pilots were presented with two engine failures during takeoff.

As a control, pilots were presented with the engine failure on takeoff presented during airline training and testing. This event is designed to practice two important skills: (a) making the decision to abort or continue and (b) maintaining control of the airplane as the flight continues. If the engine failure occurs before reaching a critical speed (known as V_1), the pilot flying may abort the takeoff. If the failure occurs at or

beyond V_1, the pilot flying must continue the takeoff. If the takeoff is continued, directional control of the airplane must be maintained as a smooth liftoff is accomplished. In our experiment, engine failure happened 3 knots beyond V_1; hence, we were looking to see that pilots continued the takeoff. The V_1 cut is required to be practiced and tested during every airline training event, during one of the takeoffs performed during the training session. However, the V_1 cut is rarely given during the first takeoff of the training session.

As the treatment condition, we gave half of the 18 pilots the V_1 cut event during the first takeoff of the session.

Procedure

Prior to flying the simulator, pilots met with the experimenters in a briefing room. Pilots were told that the upcoming simulation session would last about 2 hours and that the session would be spent practicing some of the events that are practiced during airline training and testing, along with variations on those events. Pilots were told that no abnormal events of an entirely different nature such as structural failures or fires would be introduced. Pilots were told that the purpose of the study was to evaluate training and testing methods, not the skills of individual pilots.

During the simulator session, each pilot sat in his or her respective seat: Captains occupied the left seat, first officers occupied the right seat. Each pilot was asked to do all of the flying but could delegate tasks to a confederate copilot who occupied the remaining seat. Pilots were told that the confederate pilot would not offer the help and advice that a copilot would during an actual flight. The copilot would always comply with any requests made by the pilot flying but would not provide help or information unless asked to do so. The copilot was the same for each session and is currently employed as a B747-400 pilot at a U.S. carrier.

A primary goal in designing our study was to maintain, to the extent possible, the look and feel of a realistic flight operation and recurrent training session to ensure that we were observing airline pilots performing as they do in their jobs. Thus, the seven abnormal events detailed previously were presented to pilots during three flight legs that contained a takeoff, departure, arrival, and landing. To avoid the "start/stop" nature of conventional experiments and preserve the flow of a real flight, the three legs contained other recurrent training events, both normal and abnormal, that are not reported here. All three legs began and ended at the same airport. To avoid familiarity with the airport and associated terminal area we chose an airport (JFK) to which most pilots in our sample had never flown. A few pilots had flown to JFK for the last time more than 10 years ago (while working for a previous employer).

Given the number of possible orderings of the seven abnormal events and only 18 participants, we used two simple presentation orders. Pilots who received the first presentation order experienced the unexpected stalls and wind shear events before the expected versions of these events and saw the V_1 cut event later in the session. Pilots who received the second order experienced a V_1 cut on the first takeoff of the session and were then presented with the expected versions of the stall and wind shear events prior to seeing the unexpected stalls and wind shear events. The extent to which pilots would become savvy to the fact that we were presenting abnormal events in an unexpected fashion would likely reduce the unexpectedness and diminish our results.

A cockpit voice recorder captured all verbal activity in the cockpit.

At the conclusion of the simulator session all pilots were debriefed and told the basic purpose of the study.

RESULTS AND DISCUSSION

Pilots' responses to the different versions of the three abnormal events are analyzed separately.

Aerodynamic Stalls

In comparing pilots' responses to the aerodynamic stalls, we were interested in the time it took pilots to respond to the stall by applying maximum power.

Throttle response times. The box plots in Figure 1 depict the average time that pilots took to apply maximum power after the onset of each of the three stall events, along with the second and third quartiles and the minimum and
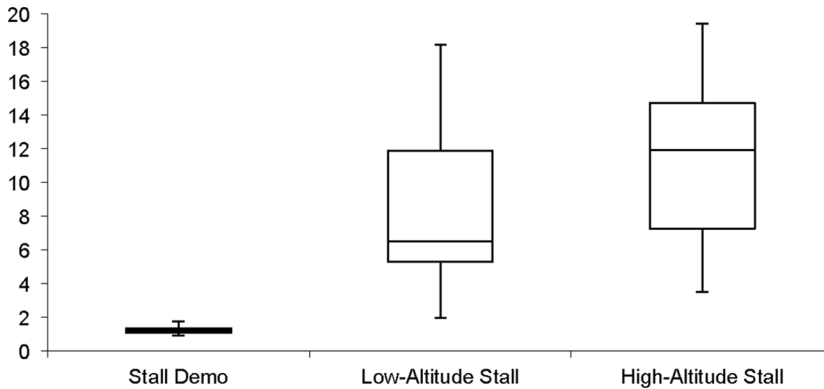
Figure 1. Delay in responding with maximum power during three stall events.

maximum time. Since all 18 pilots experienced all three types of stalls, the data presented in Figure 1 represent a within-subjects comparison.

Note that the times for the stall demonstration practiced in airline training were quite short ($M = 1.33$ seconds) and showed very little variability between pilots ($SD = 0.55$). But when the low- and high-altitude stalls were presented with no warning, response times were significantly longer and more variable. For the low-altitude stall, response times ranged between 1.9 and 18.2 seconds ($M = 8.36$ seconds) and exhibited much greater variability ($SD = 4.9$), with 2 pilots never applying maximum power. For the high-altitude stall, times ranged between 3.5 and 19.4 seconds ($M = 11.39$, $SD = 5.05$), with 6 of the 18 pilots never applying full power. A repeated measures ANOVA yielded a significant main effect: $F(2, 18) = 10.57$, $p < .001$. Post hoc tests showed a significant difference between the stall demonstration and low- and high-altitude stalls ($p < .05$), even with 6 of the 18 data points missing from the high-altitude stall condition (listwise deletion was used to handle the missing data; Little & Rubin, 1987).

There was no significant correlation between total flight experience and time to respond to the stall.

Voice recorder. A review of the cockpit voice transcript showed that of the seven pilots who delayed more than 10 seconds in responding to the low-altitude unexpected stall, five pilots specifically asked the confederate pilot if he knew the nature of the problem prior to initiating a throttle response.

Summary. These results suggest that when presented unexpectedly, pilots struggled to recognize the aerodynamic stall and delayed when responding with the familiar stall recovery procedure. Although the airplane experienced a stall buffet, a stick shaker warning, and assumed a pitch-down attitude while the airspeed had precipitously dropped, pilots seemed unsure about what was going on, and in most cases, recovery was considerably delayed. The results suggest that pilots' ability to recognize a stall may be closely tied to the behaviors exhibited by the airplane after a series of known stall-inducing steps are deliberately taken.

Low-Level Wind Shear

We were interested in four aspects of pilots' responses to the two low-level wind shear events: (a) the time it took pilots to respond with maximum power, (b) the total altitude lost, (c) the average pitch attitude pilots used during the maneuver, and (d) whether or not pilots raised the landing gear or flaps while still in the wind shear condition.

Throttle response times. The mean response times for the two wind shear conditions were $M = 7.7$, $SD = 2.2$, for expected and $M = 9.8$, $SD = 4.6$, for unexpected and are shown by the box plots in Figure 2. Again, all 18 pilots experienced an expected and unexpected wind shear
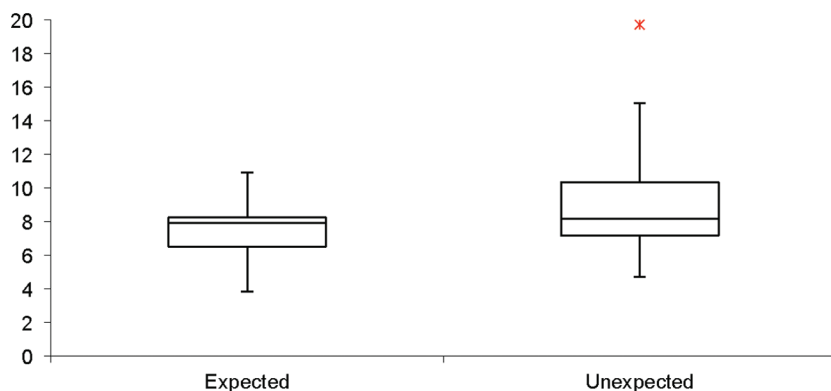
Figure 2. Delay in responding with maximum power during two wind shear events.
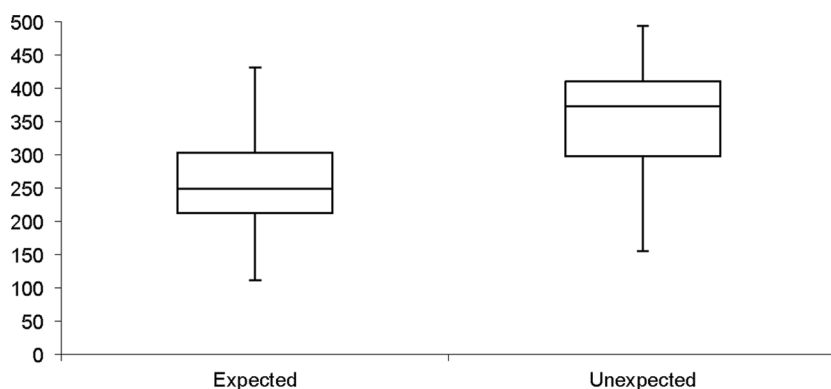

Figure 3. Altitude lost during two low-level wind shear encounters.

encounter, thus the data presented in Figure 2 represent a within-subjects comparison.

Three pilots did not advance the throttles to their maximum setting in the expected condition, while two pilots did not in the unexpected condition. A paired *t* test comparing pilots' responses to the two wind shear events revealed no significant difference.

Altitude lost. Figure 3 shows a comparison of the altitude lost by each pilot during the two low-level wind shear events.

In the expected condition, pilots lost an average of 266 feet ($SD = 83$), with a minimum of 112 feet and a maximum of 431 feet. In the unexpected condition, pilots lost an average of 353 feet ($SD = 91$), with a minimum of 155 feet and a maximum of 494 feet. There was a significant difference between the two conditions: $F(1, 18) = 8.96, p < .01$.

Pitch attitude. Average pitch was 10.03 degrees ($SD = 1.63$) in the expected condition and 9.56 ($SD = 1.47$) in the unexpected condition. The difference between the two conditions was not significant.

Configuration changes. In the expected condition, 2 of the 18 pilots changed the configuration of the airplane before the wind shear conditions had been escaped. In the unexpected condition, 10 of the 18 pilots changed the configuration of the airplane, a difference that is unlikely attributable to chance: $\chi^2(2) = 32.0, p < .01$.

Voice recorder. A review of the cockpit voice transcript showed that none of the 18 pilots mentioned wind shear when the wind shear was presented unexpectedly.

Summary. When presented with the expected wind shear encounter, pilots responded with the

standard wind shear escape procedure. When the event was presented unexpectedly, instead of leaving the airplane configured for a wind shear escape maneuver, 10 pilots changed the configuration of the airplane as is done during a simple go-around procedure. These actions resulted in a significant difference in the amount of altitude lost during the wind shear encounter. The results suggest that pilots' trained response to wind shear events may rely on cues received from weather broadcasts and on an automated cockpit system that announces the wind shear event.

There was no correlation between total flight experience and performance on any of the measures taken for the wind shear event.

Engine Failure on Takeoff

We were interested in two aspects of pilots' response to the engine failure on takeoff event: (a) making the correct go/no-go decision and (b) lateral control of the airplane throughout the maneuver.

Go/no-go decision. All of the nine pilots who experienced the V_1 cut on a latter takeoff during the experiment correctly continued the takeoff. Of the nine pilots who experienced the V_1 cut on the first takeoff of the session, two pilots (both captains) aborted the takeoff after V_1 had been reached.

Voice recorder. Both pilots who aborted the takeoff after V_1 acknowledged the verbal announcement that the critical speed (V_1) had been passed and subsequently removed their hand from the throttle levers. Both pilots who aborted later stated that this was contrary to what they had intended to do. Other pilots who correctly performed the maneuver indicated that they had been taken by surprise by the event.

Lateral control of airplane. There were no significant differences in the distance that each pilot drifted from the centerline of the runway during the engine failure event between pilots who experienced the engine failure during the first or during a subsequent takeoff: $M = 67$ feet ($SD = 59$) upon reaching an altitude of 35 feet above ground level.

Summary. Of the 18 pilots, 16 performed the maneuver well, and 2 pilots aborted the takeoff after the critical takeoff speed.

SUMMARY AND CONCLUSION

We presented a sample of 18 airline pilots with three kinds of abnormal events. When these events were presented in the familiar ways that they are taught, practiced, and tested during airline training, pilots' responses were consistent with accepted standards and varied little from pilot to pilot. When presented with these same events under less predictable circumstances, as they would surely be encountered during a real flight, pilots' responses frequently differed from accepted standards and showed greater variability. Our control conditions demonstrate that pilots' abilities to respond to the "schoolhouse" versions of each abnormal event were in fine fettle. The problems that arose when the abnormal events were presented outside of the familiar contexts used in training demonstrate a failure of these skills to generalize to other situations. This result raises the question of to what extent our training prepares pilots to respond to abnormal events in all the forms in which they might manifest themselves or only to a single example of each abnormal event. That these effects were seen in such a small sample of pilots suggests that the problems described by a few accident reports are likely more prevalent than we may have previously suspected.

While we made no attempt to directly measure the cognitive processes used by pilots while they worked, the data provide us with at least informal clues about where the breakdowns we observed may have occurred. In response to the aerodynamic stalls and wind shear encounters, pilots provided verbal evidence that they struggled with their recognition of these events. Comments made during the stall events suggest that pilots were simply puzzled by what they saw. When responding to the unexpected wind shear encounter, that the majority of pilots committed the same procedural misstep again suggested a simple miscategorization of the situation. Comments made by pilots who aborted a takeoff after a critical

takeoff speed allude to a momentary state of confusion associated with the psychological state of surprise or startle.

The performance we observed when abnormal events were presented outside of their familiar context naturally draws our attention to the characteristics of expertise. A first hallmark of expertise is the speed at which experts work (Glaser & Chi, 1988). Unlike novices who must painstakingly attempt to understand a problem situation and devise a solution from first principles, experts seem to possess an armory of ready-to-use, highly practiced associations between situations and solutions. When confronted with a familiar problem, experts often quickly recognize it and carry out a solution in a way that seems to require little conscious thought or effort (Klein, 1998). A second hallmark of expertise is that unlike novices who are easily tripped up when subtle variations on a problem or its context are introduced, experts often seem unfazed by these manipulations (Gentner, 1988). Through their experience of encountering problems in different ways, experts appear to develop a more robust understanding that allows them to recognize and respond to variations on a problem situation (Taatgen et al., 2008). Where novices are derailed, discombobulated, or taken by surprise when problems are presented under novel circumstances, experts characteristically perform as if they have "been there and done that." The performances we observed suggest that pilots are not being trained to the level of "expert" when responding to abnormal events.

Potential Solutions

Our results suggest four ways in which training and testing for abnormal events might be improved.

Change it up. The most important step suggested by our results is to abandon the idea of practicing and testing abnormal events in the same way every time. A more complete treatment of these events would allow pilots to practice recognizing them and choosing and recalling the appropriate response when presented with different forms in which each event might naturally present itself.

Train for surprise. Although skill and experience are known to reduce the occurrence of surprise (Merk, 2009), it is likely impossible to completely eliminate the element of surprise from unexpected events. Several researchers have argued that performing in surprise situations represents an additional competence area that requires special practice focused on pilots' attentional behavior and sensemaking in surprise situations (Hilscher & Kochan, 2005).

Turn off the automation. For abnormal events in which an automated system provides assistance in recognizing the event, pilots would likely benefit from exposure to these events when the automated system is inoperative to ensure that pilots are able to recognize the situation itself, rather simply respond to an alert (Wiener, 1985).

Reevaluate testing practices. Certain kinds of testing are known to encourage the use of questionable techniques such as "teaching to the test" (Bushweller, 1997; Popham, 2001; Volante, 2004) in which the content of training can be effectively "dumbed down" as time devoted to critical thinking gives way to the drill and practice of specific questions (Herman, 1992; Sacks, 2000). Training programs might consider randomizing the skills that are tested or moving some types of training outside of the realm of "jeopardy" (i.e., not formally testing them). Trainers should be aware that memorize-and-test practices not only result in poor learning but can also present the illusion that real learning has taken place when in fact it has not (Shepard, 2000; Smith & Fey, 2000).

A natural objection to these proposals is that they increase the length and cost of training. A future study might consider the question of how much practice pilots might need to broaden their skills for abnormal events. In one case, pilots might need repeated practice with each type of event. In another case, since pilots understand the basics of each of these events, pilots may be able to quickly break the mold of the memorized events they see in training and widen the scope of their skills with very little practice or possibly in ways that do not require the use of expensive simulators.

KEY POINTS

- The responses to abnormal in-flight events learned and practiced during airline training may not generalize well to more naturalistic settings.
- While existing training practices facilitate pilots' ability to respond to abnormal events, they may fail to develop pilots' ability to recognize them and recall the appropriate responses.

REFERENCES

Bushweller, K. (1997). Teaching to the test. *American School Board Journal*, *184*(9), 20–25.

Gentner, D. R. (1988). Expertise in typewriting. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 1–21). Hillsdale, NJ: Lawrence Erlbaum Associates.

Glaser, R., & Chi, M. T. H. (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hancock, P. A., & Szalma, J. L. (2008). *Performance under stress*. Burlington, VT: Ashgate Publishing Company.

Healy, A. F., Schneider, V. I., & Bourne, L. E., Jr. (2012). Empirically valid principles of training. In A. F. Healy & L. E. Bourne (Eds.), *Training cognition: Optimizing efficiency, durability, and generalizability* (pp. 13–39). New York, NY: Psychology Press.

Herman, J. L. (1992). What research tells us about good assessment. *Educational Leadership*, *49*(8), 74–78.

Hilscher, M., & Kochan, J. A. (2005, September). *From the couch to the cockpit: The application of psychological practices to upset recovery training*. Paper presented at the 3rd Annual Meeting of the Society for Human Performance in Extreme Environments, Orlando, FL.

Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science*, *8*, 255–273.

Klein, G. A. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.

Meredith, G. P. (1927). Consciousness of method as a means of transfer of training. *Forum of Education*, *5*, 37–45.

Merk, R. J. (2009). *A computational model on surprise and its effects on agent behavior in simulated environments* (Technical Report NLR-TP-2009-637). Amsterdam, Netherlands: National Aerospace Laboratory.

NTSB. (1995). *Aircraft accident report: Flight into terrain during missed approach, USAir Flight 1016, Charlotte, North Carolina, July 2, 1994*. Washington, DC: Author.

NTSB. (2010a). *Aircraft accident report: Loss of control on approach, Colgan Air, Inc. Clarence Center, New York, February 12, 2009*. Washington, DC: Author.

NTSB. (2010b). *Aircraft accident report: Runway overrun during rejected takeoff, Global Exec Aviation, Columbia, South Carolina, September 19, 2008*. Washington, DC: Author.

Popham, W. J. (2001). Teaching to the test. *Educational Leadership*, *58*(6), 16–20.

Sacks, P. (2000). Predictable losers in testing schemes. *School Administrator*, *57*(11), 6–9.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Smith, M. L., & Fey, P. (2000). Validity and accountability of high-stakes testing. *Journal of Teacher Education*, *51*, 334–344.

Taatgen, N. A., Huss, D., Dickison, D., & Anderson, J. R. (2008). The acquisition of robust and flexible cognitive skills. *Journal of Experimental Psychology: General*, *137*, 548–565.

Volante, L. (2004). Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*, *35*. Retrieved from http://umanitoba.ca/publications/cjeap/articles/volante.html

Wiener, E. L. (1985). *Human factors of cockpit automation: A field study of flight crew transition* (NASA Contractor Report NASA-CR-177333). Moffett Field, CA: NASA Ames Research Center.

Woodrow, H. (1927). The effect of the type of training upon transference. *Journal of Educational Psychology*, *18*, 159–172.

Stephen M. Casner is a research psychologist at the NASA Ames Research Center. He received his PhD from the Intelligent Systems Program at the University of Pittsburgh in 1990. Steve holds an airline transport pilot certificate with type ratings in the Boeing 737 and Airbus A320.

Richard W. Geven is a research associate with the San José State University Research Foundation. Richard holds an airline transport pilot certificate with Boeing 757 and Dassault Falcon 2000 type ratings and is currently a captain for a corporate air charter company. Richard holds a master's degree in jazz performance from the Graduate School for the Arts in Arnhem, Netherlands.

Kent T. Williams is a Boeing 747-400 first officer for a major U.S. airline and a part-time yoga instructor. Kent holds a BS in aviation management with a minor in human factors from Southern Illinois University.