FEATURE

Failure Predicts Success:

Professional Ethical Decision-Making in Aviation Simulators

Deonna D. Neal, eSchool of Graduate Professional Military Education
William H. Rhodes, Aerworthy Consulting, LLC

ABSTRACT

This research investigated the ongoing problem of pilot-induced mishaps from the perspective of professional ethics. The research relied heavily upon precedent work in philosophical virtue theory and moral psychology, including MacIntyre (1984) and Rest, et al, (1994). Anonymous field surveys were used to collect samples of behavior judged by SME's as likely to induce or preclude an aircraft mishap. These observations were reduced to a Behaviorally Anchored Rating Scale (BARS) diagnostic and to construct simulator scenarios. Participants in the simulator phase were entered into a 3 x 2 pre-test / post-test experimental design. The scenarios offered participants opportunities to display relevant behaviors and experience the resulting session outcome (safe landing at an airport or other). Participants were randomly assigned to one of three intervention groups (control, FAA, and experimental). Experimenters were kept blind to group assignment. Diagnostic scores proved predictive of session outcome. No significant difference in pre- to post-test improvement was observed between experimental groups.

Deonna D. Neal, PhD, is an Associate Professor of Leadership and Ethics, eSchool of Graduate Professional Military Education, and Adjunct Professor of Leadership and Warfighting, Air War College, Air University, Maxwell AFB, Alabama. Dr. Neal's research interests include professional ethics and the ethics of war. Dr. Neal is a former Air Force officer, general aviation pilot, and FAA certificated glider flight instructor.

Bill Rhodes, Ph.D. is principal investigator for Aerworthy Consulting, LLC, specializing in innovative results-oriented R&D in moral psychology and professional ethical development. Clients have included Avemco Insurance Company, Cirrus Aircraft, the Defense Advanced Research Projects Agency (DARPA), the TBM Owners and Pilots Association and C^2 Technologies, Inc. Dr Rhodes has taught and researched in applied ethics for nearly 30 years. He frequently lectures in ethics, mishap prevention, and professional development. His work has been featured in *Flying Magazine, NAFI Mentor,* and *General Aviation News*. Author of *An Introduction to Military Ethics: A Reference Handbook*, he is a retired military officer, former professor at the Air Force Academy and former president of the International Society for Military Ethics. Dr. Rhodes is an active pilot and aircraft owner.

Strongly significant (X2 = .007219) pre- to post-test improvement was observed in those pilots suffering a mishap in the pre-test, regardless of experimental group.

General aviation mishaps in the United States claim an average of about 500 lives annually. This statistic has remained constant over the last decade and shows no signs of improving. (See Table 1). The persistence of this accident rate is somewhat surprising, given the fact that there have been significant developments in the availability of onboard weather, GPS navigation units, as well as the introduction of aircraft parachutes over this same time.[1] The main cause (of at least 70%-80%) of general aviation accidents is pilot error. Completely satisfactory causal accounts of these "errors," however, are difficult to find. The Federal Aviation Administration (FAA) and Department of Defense (DoD) currently use the Human Factors Accident Classification System (HFACS) to analyze and describe the cause of accidents.[2] Arguably, the HFACS fails to capture some of the more nuanced dimensions of human behavior, to include the values that underlie and motivate behavior.

For example, the cause of a pilot-induced mishap may be classified as "pilot's failure to recover from an unusual attitude." While that may very well be the final (failed) action of a pilot in the mishap event, there are often antecedent events that may offer more insight into how the mishap flight evolved in the first place. For example, perhaps a pilot watched an airshow and decided that he would try to roll his airplane, without having received any aerobatic training. While at face value, the cause of the crash is a "failure to recover from an unusual attitude continued VFR flight into IMC," at least part of the root cause of the accident lies elsewhere, namely, in the pilot's failure to keep priorities straight, or perhaps even to perceive the risk involved in such a maneuver for an unqualified pilot. Unfortunately, accident reports made available to the public rarely offer the full context of the events leading to a mishap. Furthermore, it is impossible to interview the dead pilot(s) to find out what actually happened.

Mitigating Operator-Induced Mishaps (M2) sought to develop a research protocol that would more fully investigate the causes of pilot-induced mishaps and the values that underlay pilot performance. The overarching thesis of the research is that pilot-induced mishaps result more from failures of professional ethical decision-making rather than from basic "stick and rudder skills." Hence, the research protocol used a professional ethics model, previously validated in medical ethics (Bebeau 2006), and applied it to aviation mishap analysis.

The overarching thesis of the research is that pilot-induced mishaps result more from failures of professional ethical decision-making rather than from basic "stick and rudder skills."

Theoretical Background

The theoretical framework deployed by M2 is a hybrid that combines a psychological model of ethical decision-making with a professional ethical model grounded in philosophical virtue theory.

The Four-Component Model

The project uses an adapted version of the University of Minnesota's "Four Component Model" (FCM) of ethical development for diagnostics and scoring. The social science fundamentals underlying the project are well

1 Cirrus Aircraft introduced a ballistic recovery system (BRS) also known as the Cirrus Airframe Parachute System (CAPS) in 2002.

2 The original framework (called the *Taxonomy of Unsafe Operations*) was developed using over 300 Naval aviation accidents obtained from the U.S. Naval Safety Center (Shappell & Wiegmann, 1997). The original taxonomy has since been refined using input and data from other military (U.S. Army Safety Center and the U.S. Air Force Safety Center) and civilian organizations (National Transportation Safety Board and the Federal Aviation Administration). The result was the development of the Human Factors Analysis and Classification System (HFACS).

established and work from the Center for the Study of Ethical Development at the University of Minnesota[3] and professional ethics education in other fields form the basis for the research (Rest & Narvez 1994, Rest & Narvez 1999, Beabau & Monson 2008) According to Rest et. al, the four components of ethical decision-making are perception, judgment, commitment and competence. For example, a person needs to first *perceive* that there is an ethical issue at stake; *deliberate* as to the best course of action to resolve the problem; *commit* to following through on the chosen course of action; and be *competent* to carry out the course of action. The FCM was deployed successfully in the field of dentistry. The results of that research showed that when dental students were introduced to the FCM during their training they had a lower rate of malpractice when tracked longitudinally (Bebeau 2006).

M2 adapted the FCM to the aviation domain. The four components, as modified are: Perception: pilot sensitivity to and detection of factors important to effective decision-making; Judgment—effective decision-making, especially in ambiguous situations; Commitment—the ability to carry out good decisions in the face of temptation to do otherwise; Competence—the skills to execute decisions reliably.

Virtue Theory and Professional Ethics

While the FCM provided the basic heuristic for categorizing ethical behaviors, M2 enfolded the FCM into an overarching virtue ethics model. The virtue ethics model is basically Aristotelian, as articulated by Alasdair MacIntyre (1984). A virtue theoretic approach insists that every activity has a *goal* or *telos*. Likewise every *craft* has an overall *goal* as well.[4]

With respect to aviation, the overall goal of the craft of aviation is judged to be a safe landing, or "on the ground and OK." While it is true that airplanes can be used in a variety of ways, e.g. as a means of transportation, for aerobatic demonstrations, for pleasure, etc., no one would judge any pilot to have met the goal of the craft of aviation if he failed to land the plane successfully, regardless of the particular use of the aircraft at the time.[5] In other words, if a safe landing has not been achieved, it is impossible to evaluate that flight as "good," even if all the other actions during the flight were executed perfectly. When mechanical failures arise, or if some other condition not attributed to pilot error occurs, virtue ethics would judge the pilot virtuous if the pilot successfully negotiated a landing without injury to himself and others, or if he undertook a course of action to mitigate injury to others, either in the plane, or on the ground, as much as possible. An example of such a case might be a pilot who has an engine failure during flight and steers his aircraft away from houses and populated areas. The pilot might die in the crash, but his actions minimized the loss of life to others.

If a safe landing has not been achieved, it is impossible to evaluate that flight as "good," even if all the other actions during the flight were executed perfectly.

M2 used the FCM and virtue theory as the basic theoretical model to frame the problem of pilot-induced error. The FCM was used to form the basis of a diagnostic tool to evaluate pilot behaviors in the area of perception, judgment, commitment, and competence, and used the virtue theoretic approach to establish the

3 The Center is now located at the University of Alabama.

4 For example, the goal of using a hammer is to drive a nail, and the overall goal of the craft of carpentry is to build or repair something using wood. Behaviors are valued as "good" or "bad" in accordance with how well they serve the function of the craft. A craftsman is considered "virtuous" to the extent that his behaviors are functionally oriented and ordered to achieving the overall goal of the craft.

5 The use of aircraft for military purposes might prove to be an exception. However, the use of aircraft in war is purely instrumental and is subsumed into the larger craft of warfare, whose goal is victory.

desired outcome of a flight, i.e. mishap or non-mishap.

Research Hypotheses

In light of the theoretical model used above as well as new work in cognitive neuroscience, M2 tested four research hypotheses:

- H1: *Pilot-induced simulator mishap rates are negatively correlated to scores in the 4 components.*
- H2: *Simulator performance scores are positively correlated to scores in the 4 components.*
- H3*: M2 educational intervention will improve scores in components one, two and three on the post-test.*
- H4: *M2 educational intervention will improve simulator performance scores on the post-test.*

Method

The research methodology consisted of three phases: 1) survey work; 2) simulator scenario design; and 3) data collection.

Phase I: Survey Work

Using the categories of perception, judgment, and commitment from the FCM, a survey was designed and distributed to flight crew subject matter experts (SME) around the country using Survey Monkey. Pilots and non-pilot flight crew participants, who had at least 1,500 hours of total aircraft time, were asked to provide basic demographic data and to answer open-ended questions which pertained to behavior that they had observed in the cockpit that correlated to excellent, average and poor examples of flight crew perception, judgment, and commitment. Survey respondents were also asked to list one trait of the pilots who scared them the most and one trait of the pilots that they trusted the most.

The first survey had 119 respondents who provided over 430 discrete pilot behaviors. After this first round of survey data was collected and sorted, we consolidated the 430 behaviors (some responses were repetitive or irrelevant to the question asked) to 213 behaviors. We then sent this more refined data to four independent SMEs, who validated the initial behavioral component sorting. Following this step, we then sent this consolidated and sorted set of 214 behaviors to "Super Subject Matter Experts" (SSMEs). SSMEs had to be pilots or flight crew with a flight instructor rating who had at least 3000 hours of pilot time and 700 hours of dual given. The SSMEs scored each of the 213 responses on a scale of 1 to 5, where a "1" designated a behavior as "least likely to cause a mishap" and a "5" designated a behavior as "most likely to cause a mishap."

After the SSMEs scaled the individual behaviors, a factor and item analysis was performed on the 213 behaviors. At the end of the factor and item analysis, 16 behaviors were identified as being the most indicative of likely mishap and not-likely mishap behavior. These 16 behaviors were then used to form a Behaviorally Anchored Rating Scale (BARS), which would serve as the basic diagnostic and evaluative tool for pilot performance during the simulator scenarios.

> *16 behaviors were identified as being the most indicative of likely mishap and not-likely mishap behavior.*

Phase II: Simulator Scenario Development

The research used FRASCA T-6A Texan II simulators located at the United States Air Force Academy. Due to the anticipated participant demographic and experience level, the simulator was modified to perform like a high performance single-engine airplane, similar to that of a Cessna 210 or a Bonanza, as opposed to a high-performance military turbine trainer. This was accomplished by limiting the power output of the simulator, as well as by modifying the simulator instrument panel to look more like a general aviation aircraft. For example, redundant instruments as well as a number of warning annunciators that would not be found in a general aviation aircraft were covered up.

The simulator scenarios were designed using the NTSB accident record as well as mishap analysis gained from the aviation insurance industry.[6] Two Visual Flight Rules (VFR) scenarios were designed. Each scenario required the participant to act as pilot-in-command (PIC) while carrying a passenger (played by members of the research team). Since the accident record shows that "continued VFR flight into IMC" and "fuel exhaustion" mishaps continue to plague the general aviation community, one scenario of each kind was designed for the experiment.

The simulator scenario design was limited by the visuals in the simulator, which provided a graphical range from about 10 miles east, 20 miles west, and about 35 miles north and south of Colorado Springs Municipal Airport (KCOS). The complete runway environments for the Air Force Academy (KAFF) and KCOS, as well as the surrounding areas were available to use for the scenarios. However, due to the limited range of the visual graphics, the scenarios had to be designed to begin during the "enroute" phase of flight.

Two researchers enacted tightly scripted roles in each scenario, playing the roles of Air Traffic Control (ATC) or passenger. In Scenario A one member of the research team played the role of ATC and managed the simulator event inputs, while the other member played the role of a passenger. During Scenario B, the research team switched roles, i.e. the passenger in Scenario A became the Air Traffic Controller for scenario B, and vice versa.

During the simulator scenario development phase, the research team also designed a complete "standard weather briefing" which was provided to each participant to use for pre-flight planning. This briefing package was based on the content of the official services that would be provided to pilots by a Flight Service Station (FSS) during a real flight. In order to make the scenarios realistic, the researchers downloaded actual weather data for a discrete time period during two typical weather environments that could be found in Colorado. The pre-flight briefing package included radar, satellite, and surface analysis reports as well as standard aviation weather data reports, such as Terminal Area Forecasts (TAFs), Meteorological Reports (METARs), Graphical AIRMETs and SIGMETs, Area Forecasts (FA), and Winds Aloft information. The member of the research team who was acting as the air traffic controller adjusted the simulator visual weather environment during the scenario to correlate to the weather information that was provided.

Phase III: Data Collection

The research protocol used a 3 x 2 pre-test/post-test design. Scenario A was a VFR into IMC flight and Scenario B was a fuel leak incident. The three different intervention groups were: 1) Control Group; 2) Federal Aviation Administration Aeronautical Decision-Making Group (FAA ADM); and 3) Professional-Ethics Experimental Group (aka Trustworthy Pilot Group). The number of participants needed to yield the requirements for statistical significance was determined to be 108. This allowed for 36 participants in each intervention group. Furthermore, the scenarios were counterbalanced. For example, 18 participants in the control group flew scenario A first, followed by scenario B; the other 18 participants in the control group flew scenario B first followed by scenario A. This counter-balancing was designed to washout any discrepancies in scenario difficulty. Furthermore, participants were randomly assigned to their scenarios and their intervention groups by a third party. As a result, the experimenters did not know what experimental group the participants were in until after the entire protocol

6 One of the researchers has a non-disclosure agreement with an aviation insurance carrier and was able to use that knowledge in a way to help shape the scenario design, but without violating the non-disclosure agreement.

was completed and could thus score pilot performance without bias.

Participants were recruited from the Colorado Springs and Denver Metro Area. Recruiting posters deployed at the United States Air Force Academy, at Colorado Springs and Denver Metro Area Airport Flight Schools, as well as among various aviation groups. Many flight schools and aviation groups also agreed to distribute copies of recruiting posters to their members via group lists. Additionally, much of the recruiting happened by word of mouth, as early research participants told their fellow pilots about their positive experiences of the research protocol.

Participants were randomly assigned to their scenarios and their intervention groups by a third party.

To qualify as a research participant a pilot had to: 1) Possess at least a student pilot certificate and be qualified to solo an airplane single-engine land (ASEL); 2) have completed at least three takeoffs and landings in the previous 90 days; 3) be at least 17 years of age; 4) not be pregnant and 5) be willing to consent to neurophysiological monitoring.

The research protocol required the participants to come to the Air Force Academy Air Warfare Laboratory on two different days. The first session took 2.5 hours. During this session, a participant signed an informed consent document (ICD), was given a cockpit orientation in the T-6A Texan II and given the opportunity to practice some basic maneuvers in order to gain familiarity with the simulator. After the orientation phase, the participant was asked to fly a "screener" scenario where they were asked to fly a basic VFR flight from KCOS to KPUB (Pueblo) in visual meteorological conditions (VMC). This screener scenario was designed to make sure that the pilot could fly to the FAA practical test standards (FAA 2012)[7] and handle the communication requirements with ATC.

If the participant passed the screener scenario, he was admitted into the formal phase of the research protocol. The first part of the research protocol was a pre-flight planning phase, where the pilot was seated alone in a classroom and asked to prepare for the upcoming pre-test scenario using the standard weather briefing materials provided by the researchers. The researchers also provided a "case description," which included the circumstances of the flight time of day, fuel on board, and information about the passenger. Participants were allowed as much time as they needed to prepare for the flight. No false or misleading information was given to the participants at any time. While the amount of time each participant used to do pre-flight planning was not officially recorded, the average amount of prep time taken was approximately 20 minutes. After participants completed their pre-flight planning, they were hooked up to psycho-physiological monitoring equipment.[8].

When the preflight preparation and physiological hookup was complete, participants returned to the simulator room to

7 https://www.faa.gov/training_testing/testing/test_standards/media/faa-s-8081-14b.pdf.

8 A BioPac MP 150 Data Acquisition wireless system and its accompanying software, AqKnowledge, was used to collect and record physiological data. Participants had electro dermal activity (EDA) electrodes hooked up to the *thanar* and *hypothanar* regions of the palms of their non-flying hand. This physiological measurement was designed to capture arousal in the sympathetic nervous system. Two electromyography (EMG) electrodes each were placed on the *flexor radii carporalis* and *extensor radii carporalis* muscles of the flying forearm in order to measure stress/grip strength, with a fifth electrode placed on the wrist serving as a ground. Three electrocardiogram (ECG) electrodes were placed on the participant's chest. The ECG electrodes were used to gather basic cardiac information to include heart rate, heart rate variability and vagal tone. The vagal tone measure was used as a proxy for parasympathetic nervous system activity. The ECG wireless transmitter was also attached to a chest harness to measure respiration. Additionally, there was a small camera in the cockpit that recorded facial micro expressions. The micro expression data was also used as a proxy to capture a pilot's mental states during the scenarios. The physiological data has not yet been analyzed and is not related to the non-physiological findings of this research. It is noted here to acknowledge that it was part of a participant's overall experience

begin the pre-test scenario. The researchers performed their assigned roles as ATC or the passenger. Three cameras and one microphone were in use during the simulator flight. One camera with microphone was used to record the participant's facial expression and serve as a voice recorder. A second camera was mounted on top of the simulator to record the simulator visual graphics. A third recording device was used to capture the simulator operator's board. The simulator operator's board contained airspeed, altitude, heading and course track, as well as the controller weather input and system changes. All three cameras and microphone were in simultaneous operation. An iSpy software package was used to simultaneously capture and record the participant's facial and voice data, the simulator visual graphics, and controller board. Video and audio recording of the entire flight was important for future analysis and event reconstruction.

This screener scenario was designed to make sure that the pilot could fly to the FAA practical test standards (FAA 2012)[7] and handle the communication requirements with ATC.

After a participant completed the pre-test scenario, the researcher, who played the role of passenger, immediately conducted a post flight interview, asking what was going through the participant's mind during selected events. The "non-interviewing" member of the research team, i.e. the one who had played the role of ATC during the flight, took notes on the interview and then entered those notes into a database for future analysis. After the post-flight interview was over, both researchers scored the participant's performance on the BARS. Any discrepancies in scoring between the two researchers were resolved between them and one common BARS performance score was given. The participants BARS scores were entered into a database and the paper version of the record was also maintained.

After the pre-test scenario was completed, the participant was given instructions on what to expect for the intervention assignment. The participant's post-test return date was also confirmed at this time. No training or any flight "debriefing" was done with any of the participants.

As mentioned previously, there were three experimental groups, with 36 participants randomly assigned to each group. After the participants completed their pre-test, an independent third party emailed the participants with their intervention group assignment and relevant instructions. If participants were in the control group, they had no intervention assignment and were cleared to return for their post-flight at the previously agreed upon date. If participants were in the FAA ADM group, the link to this online course was sent to them.[9] They were asked to view the course and complete the end of course quiz. After participants completed the quiz, they were asked to forward the course completion certificate to the independent third party for verification. Once the third party received the course completion certificate, the participant was cleared to return for the post-flight test. A similar procedure was used for those in the Professional Ethics Intervention Group (aka the Trustworthy Pilot group). The researchers designed the professional ethics intervention course and its course completion quiz prior to the start of the research protocol and uploaded it onto a secure link, which was sent by a third party to the participants. Like the FAA course, a participant in the Trustworthy Pilot group was asked to view the 30-minute video, take the end of course quiz, and send the quiz to the third party for verification. Once the quiz was returned, the participant was cleared for the post-flight test.

When participants returned for the post-test, they followed the same procedure that was used for the

9 https://www.faasafety.gov/gslac/ALC/course_content.aspx?pf=1&preview=true&cID=62

pre-test, except that they did not have to undergo the cockpit orientation and screener exercise again. The post-flight session took about 1.5 hours, though, again, the session length varied depending on the length of time the participant took for preflight planning. While the researchers switched roles for a participant's post-flight, the protocol was otherwise exactly the same, i.e. physiological data was collected, the post-flight interview was conducted, BARS score assigned, and information entered into the database. At the end of the post-test, the participants were informed that they had completed the research protocol and would be invited to a future seminar where the results of the study would be presented. Data collection began in November of 2013 and was completed during the first week of July 2014. Simulator trials ran Monday through Saturday, between 0800 and 2000.

Results

One hundred and sixteen (116) participants entered the study. One hundred and nine (109) participants completed the full protocol. The results of the first 108 participants were used for the data analysis. Two participants were disqualified from the study because they failed to pass the screener; two participants had to withdraw from the study due to scheduling/moving conflicts; and data was lost or incomplete on three additional participants.

Two of the four research hypothesis were supported.

- H1: *Pilot induced mishap rates are negatively correlated to scores in the four components*. Supported with $X^2 <.05$. (See Table 2 and 3)

- H2: *Simulator performance scores are positively correlated to scores in the four components*. Supported with $X^2 <.05$. (See Table 2 and 3)

- H3: *M2 Educational program will improve scores on components one, two and three*. Not supported.

- H4: *M2 educational program will improve simulator performance scores*. Not supported.

Discussion

The validation of research hypotheses H1 and H2 demonstrate the relevance of professional ethics and moral psychology in diagnosing pilot mishaps, since the BARS behaviors (components 1-3 of the FCM) were tightly correlated to simulator outcome (See Figure 2 and Figure 3). Hence, while a focus on traditional "skills" and "aeronautical knowledge," which are typically used by the aviation industry to train and evaluate pilots is certainly necessary, our research suggests that focusing on perception, judgment (deliberation), and commitment (self-discipline) may prove to be effective categories for evaluating pilot behavior as well. All of the pilots who experienced unsafe outcomes in the simulator were qualified and current, as were all of those who flew safely. The difference may be illuminated by reference to professional ethics.

All of the pilots who experienced unsafe outcomes in the simulator were qualified and current, as were all of those who flew safely. The difference may be illuminated by reference to professional ethics.

Virtue theory helps us understand that professional performance is not solely the product of technical training or skills-acquisition. Internal psychic states—perhaps most importantly what a person *cares* about—is important too. Having one's values straight, and understanding why, matters in professional ethics. In the case of aviation, caring about safety matters, and it is more than knowing how to be safe. It is acting in accordance with the value of safety that matters. It may be that the experience of caring about safety, but then experiencing an unsafe outcome, causes introspection and a subsequent reordering of behavior that increases the probability of safe outcomes.

While research hypothesis H3 and H4 were not supported, the research team is not, upon reflection, completely surprised. The initial plan for the M2 professional ethical education intervention model, i.e. Trustworthy Pilot group, was for it to be conducted in small group seminars and one on one expert/non-expert coaching. However, given the fact that the second intervention group (the FAA ADM course) was an online course, it was determined that having the Trustworthy Pilots course in the same format, i.e. an online course, was more scientifically appropriate. The concern was that the personal interaction *per se* for the experimental intervention group would prove to be a confound in the experimental design. Future research will explore the small group seminar approach for teaching professional ethics.

> *Having one's values straight, and understanding why, matters in professional ethics. In the case of aviation, caring about safety matters*

Unexpected Results

In addition to the results of H1-H, the most important research finding of M2 *is that there was a statistically significant improvement across all three intervention groups if the participant failed the pre-test* (See Figure 4). There was not a statistically significant improvement in the post-test for participants who passed the pre-test (See Figure 5). In other words, the experimental protocol itself proved to be a sort of training program. Since there were improvements across all intervention groups, the researchers concluded that the actual experience of *failing* during the pre-test scenario, proved to be the best predictor of success in the post-test (See Figure 6). The researchers believe that a participant's strong sense of identity of being a pilot, as well as knowing that other pilots performed the scenarios successfully, provided the intrinsic motivation necessary to autonomously evaluate their own performance and seek to improve on the post-test. Indeed, a large majority of the pilots who failed their pre-test spontaneously reported to the research team that they vowed to do better on the post-test and that they spent a lot of time "thinking" and "kicking themselves" and "evaluating" their performance on the pre-test.

Conclusion

M2 demonstrated that a professional ethical decision-making model can be used to design a diagnostic tool which correlates to pilot performance. A professional ethical intervention model in a 30-minute online format does not produce improved simulator outcomes or significantly improve pilot performance BARS scores. A professional ethics intervention model may prove successful when integrated into a pilot training program, if it is done in a one-on-one or small group setting. Researchers believe the language of professional ethics may prove especially powerful in helping "failed" pilots reflect upon their experience and thereby improve. Finally, the experience of failure on the pre-test proved to be the single most significant factor in predicting pilot success in the post-test.

• • •

Author Note

This research was supported by DARPA BAA grant 11-65-FP-29 awarded to Aerworthy Consulting, LLC (Dr. Bill. Rhodes) and with a visiting research fellowship awarded to Dr. Deonna Neal by the Center for Character and Leadership Development at the United States Air Force Academy.

Correspondence concerning this article should be directed to William H. Rhodes, Principal, Aerworthy Consulting, LLC, 8118 E. Greenland Rd, Franktown, CO 80116, brhodes@aerworthy.com.

References

Bebeau, Murial J. & Verna E. Monson. (2008). Guided by Theory, Grounded in Evidence: A Way Forward for Professional Ethics Education. In L. Nucci & D. Narvez, (Eds.), *Handbook on Moral and Character Education* (pp. 507-534), New York, NY: Routledge.

Bebeau, Muriel J. (2006). Evidence-based Character Development. In Nuala Kenny & Wayne Shelton, (Eds.), *Lost Virtue: Professional Character Development and Modern Medical Education* (pp. 47-86), New York, NY: Oxford University Press.

Federal Aviation Administration. (2014). Aeronautical Decision-Making for VFR Pilots. FAA Wings Safety Course. Retrieved from https://www.faasafety.gov/gslac/ALC/course_content.aspx?pf=1&preview=true&cID=62

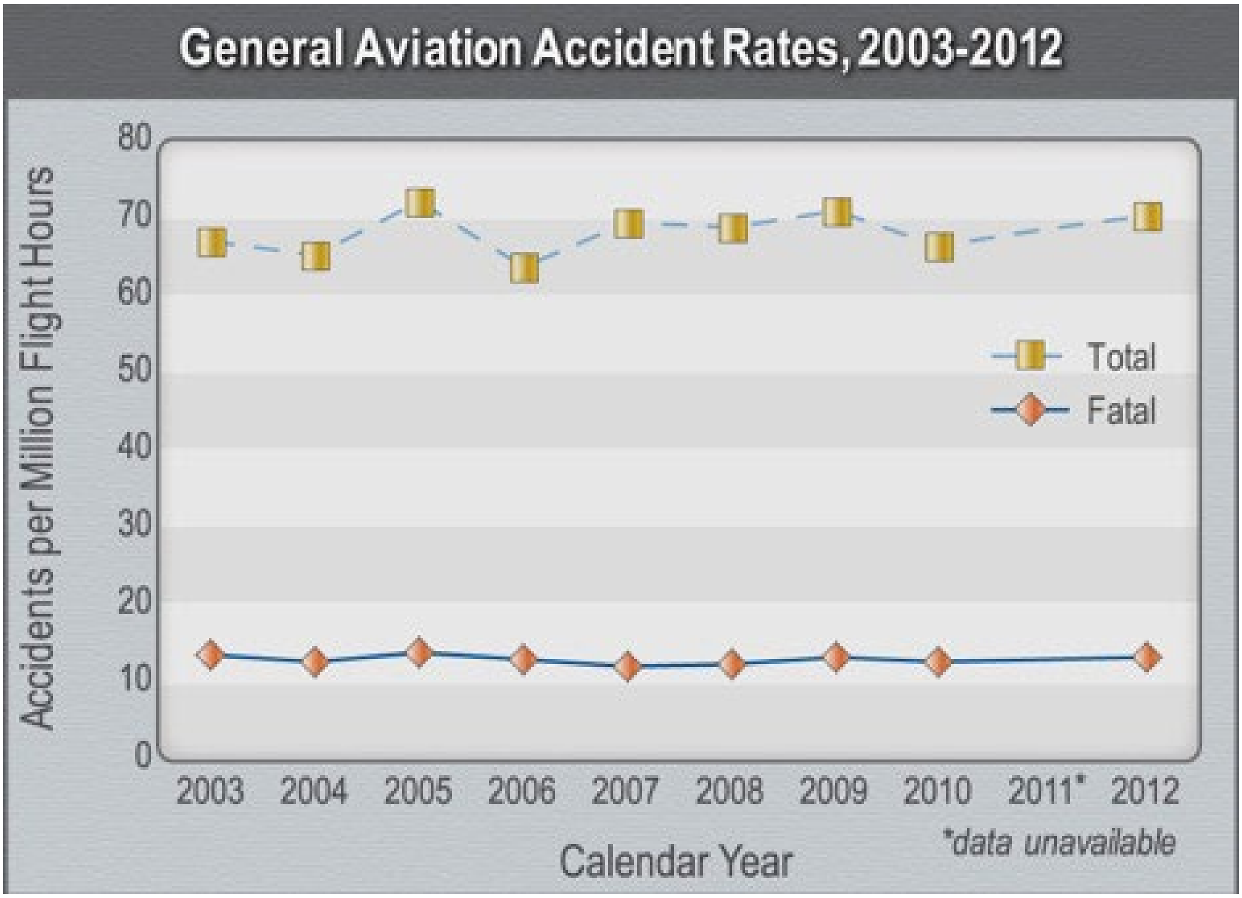
Federal Aviation Administration. (2012). Practical Test Standards: Private Pilot: Airplane Single-Engine Land. Washington D.C., Flight Standards Service. Retrieved from https://www.faa.gov/training_testing/testing/test_standards/media/faa-s-8081-14b.pdf

McIntryre, Alasdair J. (1984). *After Virtue*, (2nd ed.). Notre Dame, IN: Notre Dame University Press.

Rest, J.R. & Darcia Narvaez, Eds. (1994). *Moral Development in the Professions: Psychology and Applied Ethics*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1994.

Rest, J.R., Darcia Narvez, Muriel J. Bebeau, et. al. (1999). *Post-Conventional Moral Thinking: A Neo-Kohlbergian Approach*. Mahwah, NJ: Lawrence Earlbaum Associates.

Figure 1



http://www.ntsb.gov/investigations/data/Pages/2012%20Aviation%20Accidents%20Summary.aspx

Figure 2: Average Pre-Test BARS Scores, BARS outcome correlation (X2 <.05), N=108

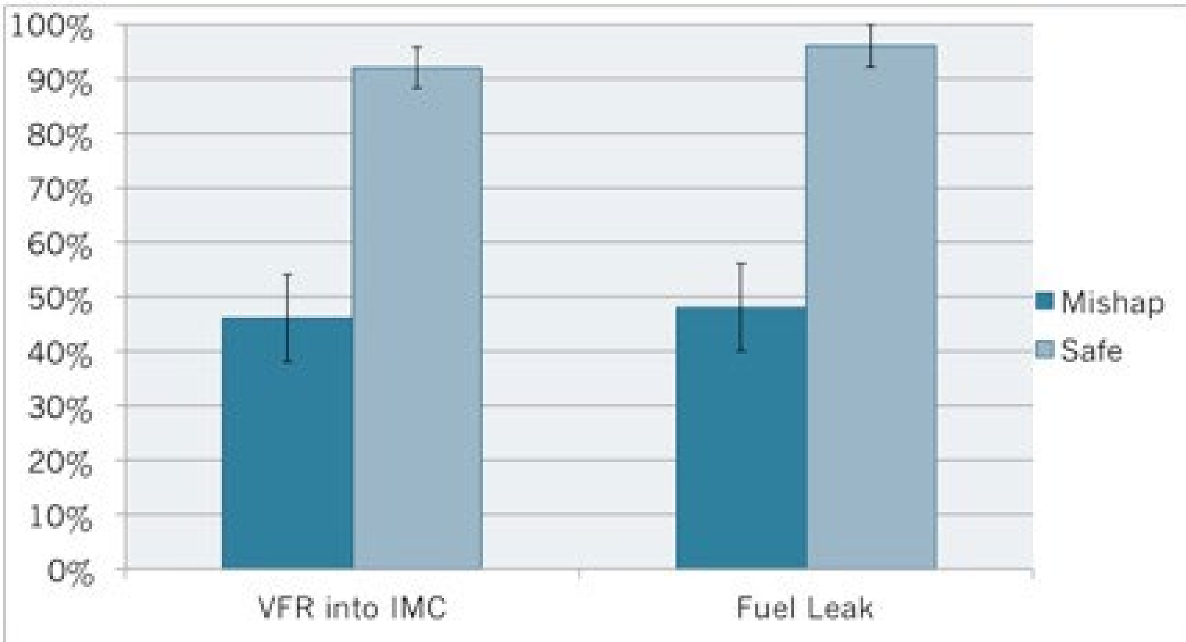


Figure 3: Average Post-Test BARS Scores, BARS outcome correlation (X2 <.05), N=108

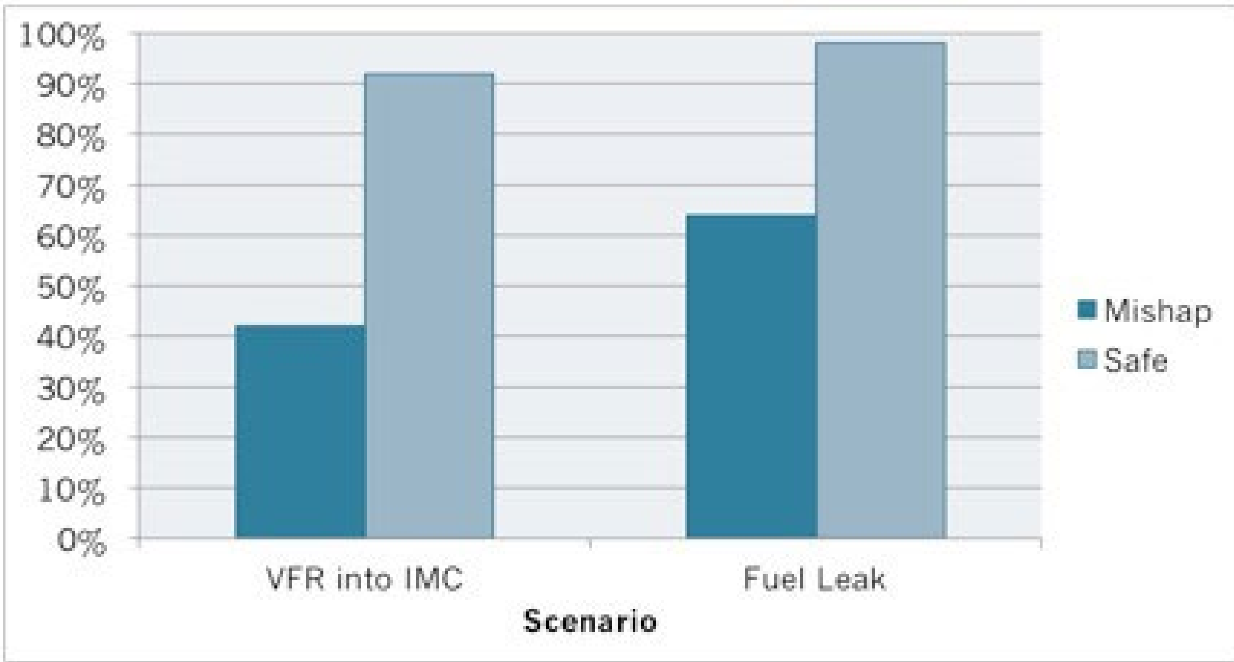


Figure 4: Post-Test Results for the Pre-Test "Mishap" Pilots, (n=68)

Figure 5: Post-Test Results for the "Safe" Pre-Test Pilots, (n=40)

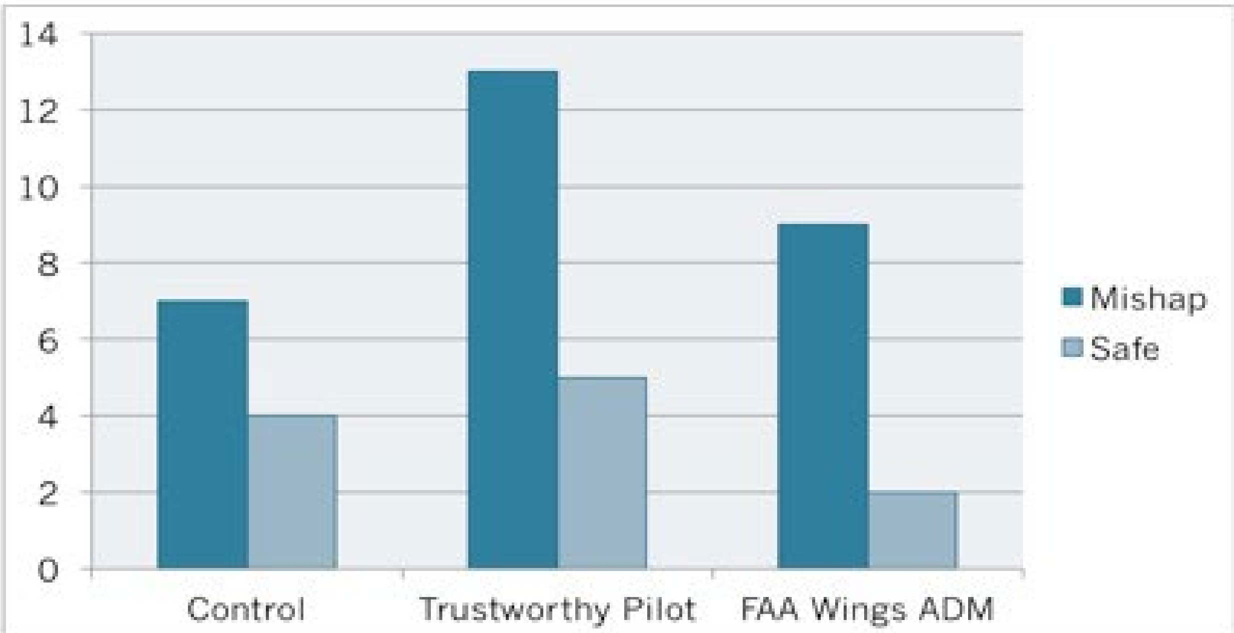


Figure 6: Summary Data for Experimental Results

	Post-Test MISHAP	Post-Test SAFE	Chi Squared
68 Pre-Test MISHAPS	27	41	
(Expected Outcomes)	(43)	(25)	.007219 (strongly significant)
40 Pre-Test SAFES	29	11	
(Expected Outcomes)	(25)	(15)	.715001 (no significant correlation)