Flight Test *Safety* Fact

Published for the Flight Test Safety Committee

Chairman's Corner *Tom Huff*

Greetings from the Flight Test Safety Committee, and welcome to the September edition of FTSF. The Committee has reviewed the feedback from the 2019 Flight Test Safety Workshop and is already deep into planning for the 2020 Workshops. We are pleased to announce that we intend to resurrect the European FTSW! Mark your calendars for 14-16 Oct 2020 to join us in London at the Royal Aeronautical Society. The theme will be Safety Risk Management in flight test. Missed the 2018 FTSW in Arlington, TX? Attended but need a refresh? Then the tutorial in London is not to be missed. We will also be in Denver, CO, 5-7 May 2020. The theme will center on Safety Promotion in flight test. Whether your company or test organization has a mature Safety Management System or not, the tutorial content and technical discussions can aid in not only implementing/improving a SMS but also boosting the safety of your flight test operations.

Speaking of SRM, the Committee recently had a lengthy email exchange regarding risk management of arguably (very) elevated risk test events (Vmca). Some excerpts from this exchange are included herewith, because it stimulates relevant discussion/debate on robust and effective SRM. As I've said publically, SRM should be a strength of testers. Sniffing out hazards and implementing mitigations should be in our DNA. If you are a fan of Operational Risk Management, then think "in-depth" ORM at the planning phase. Whatever you call it, **we should afford opportunity for test teams to do a thorough analysis of hazards and develop mitigations that reduce risk to as low as reasonably practicable**. Have you seen that statement before? How do we know we're there? Signed test plan, completed THAs, TRR, SRB? Is it clear who accepts risk on behalf of the company? CAPT Jim Wetherbee, USN (Ret), makes an important distinction in his excellent book *Controlling Risk in a Dangerous World* between managing and controlling risk. Risk [acceptance authority] managers rely on notionally calculated risk, the product of likelihood and consequence. Risk controllers–those who put their body in an aircraft–must deal with hazards real-time, with prescribed [expected] action and [discretionary] techniques for a favorable outcome. Returning to risk categorization: Consequences are relatively easy to imagine. Probabilities are an entirely different matter. How comfortable are risk accepters with subjective/probabilistic assessments? Does organic test planning process focus more on the hazard ID and mitigations over risk categorization or risk characterization (LOW/MED/HIGH)? How does your organization accommodate mitigated vs residual risk? We know there are a lot of stories out there regarding risk management, and we'd like to hear from you. *Tom Huff, Chairman*

Flight Test Safety Committee

Tom Huff, Chairman chairman@flighttestsafety.org
Society of Flight Test Engineers edir@sfte.org
Society of Experimental Test Pilots setp@setp.org
Contact *Flight Test Safety Fact*
Mark Jones Jr, Editor mark@flighttestfact.com

Letter to the Editor

Rodrigo Huete is a member of the Flight Test Safety Committee and a manager actively engaged in the oversight of flight test. Recently, a DER test pilot attempted to make his case about assessing hazards for Vmca and SHSS with an external store on a light twin-turboprop aircraft. I spoke with Rod on the phone and exchanged more than twenty emails with Committee members about the topic. Rod's email provoked a lot of discussion, and thus, we wanted to share it with you here.

Letter to the Editor ***Rod Huete***

Safety Management and planning is not black and white, and deciding where to draw lines of hazard consequences and probabilities is very subjective. It was always meant to be very subjective and dependent on gray-bearded knowledge and experience. FAA Order 4040.26 gives some examples of known risk categories (page G-1). When we wrote that, we selected those examples from experience, in particular, from Don Armstrong's and Jim Richmond's list from the Los Angeles ACO. We identified these so people who had trouble figuring out risk categories would at least have a starting point. This was prior to the NASA/FAA Flight Test Safety Database. When we started that project, we took inputs from several OEMs and asked the National Test Pilot School to populate it based on the inputs we received. If we did not have any inputs for a particular maneuver, the NTPS developed the data. Then we had a review board between NASA and the FAA (Pete LeVoci for helos, and myself and John Hed for fixed wing) to edit (or approve) the data before publication. As you could see at the bottom of each THA, we put a big caveat that this was a starting point and organizations needed to tailor the data to fit their individual programs. Nevertheless, I personally tend to favor what is in the database in lieu of better available guidance.

Some maneuvers are hazardous by their very nature, such as Vmca. In this case, the hazard is not so much losing the good engine but more about operating near the stall, in an asymmetric condition and at low altitude where even wearing a parachute would be impractical. If you want to put it in terms of the risk table, the consequence would be catastrophic and the probability would at least be occasional or even probable; remember, we are thinking worst case when we do this.

Letter to the Editor (continued)

A common problem in safety thinking is that many people fall into thinking, "I've always done it that way and nothing happened therefore we are safe". This kind of thinking is what I call the "fish in the blender" syndrome, where the fish is happy swimming in the blender but is a button push from disaster.

I have a personal bias to declare SHSS as hazardous based on experience. There are many cases where these were considered low risk yet several airplanes have had close calls and even accidents. In many cases, it is not the pilot experience that is an issue but the sudden and unperceived loss of stability (edge of the cliff which is not felt until it happens). I don't know how many times I've heard that the chief pilot or the project pilots are the ones who have these incidents/accidents, and these are people with the most experience in the airplanes they were testing. *Rod Huete*

Editor's Note: There were several emails in follow-up to Rod's conversation starter. Some of these included stories about several SHSS incident, near misses that would have been catastrophic had management not put more mitigation in place. The FTSW has also regaled attendees with similar stories. Even when test teams have applied the lessons of previous mishaps, incidents can still happen, and the appropriate mitigation provides margin to recover the aircraft in time.

1/31/2012 4040.26B
Appendix G

APPENDIX G. TYPICAL EXAMPLES OF FLIGHT TESTS AT VARIOUS RISK LEVELS

NOTE: These are typical examples only, provided here for general guidance. The actual risk category must be evaluated on a case-by-case basis and it may be different from these examples depending on actual project-specific circumstances.

HIGH RISK

- Stall characteristics:
  - Aft cg accelerated stalls with rapidly changing dynamic conditions.
  - On airplanes equipped with unproved pusher systems that are masking potential deep stalls.
  - High altitude stalls on airplanes with potential engine flameout problems.
  - With critical ice shapes.
- High speed tests above $V_{ne}/V_{MO}/M_{MO}$.
- V_{MCA} tests at low altitude; particularly dynamic V_{MCA}.
- Flight control malfunction testing during takeoff and landing phases of flight, and asymmetric deployment of roll controls at high speeds.
- Ice shape testing, especially during the takeoff phase where special procedures are required.
- Maximum energy RTOs where wheel/brake fires are a possibility.
- Actual V_1 fuel cuts for takeoff performance.
- Autopilot malfunction tests at low altitudes.
- WAT limited takeoffs with actual engine cuts.
- V_{MU} test at low thrust to weight ratios.
- V_{mcg} tests.
- Nose-wheel steering malfunction tests.
- Spin testing.
- Lateral-directional testing on aircraft that can achieve extremely large sideslip angles.
- Dynamic lateral stability testing (Dutch rolls) on airplanes that are extremely unstable under certain conditions.
- In-flight thrust reverser deployments.
- Systems installation (with unproved design aspects) where FHA has identified catastrophic events.
- Stall characteristics on Restricted Category airplanes with asymmetric wing store configurations.
- H/V envelope determination.
- Helicopter low speed testing.
- Autorotation.
- PIO Testing.
- Max Crosswind Landings.

FAA Order 4040.26B, Appendix G – Typical Examples of Flight Tests at Various Risk Levels.

Three Heuristics for Communicating Uncertainty in Flight Test

Mark Jones Jr.

The Chairman's column and Rod's letter raise many questions, but I would like to single in on the question that came to my mind as I read Rod's email: is SHSS really high risk?

Since there are two inputs into the risk assessment: hazard probability and severity. I began to wonder which would need to change. In my own thoughts, the word "change" caught me by surprise, because I wasn't sure I had an idea of where I was starting from. I did know one thing for certain: I have really strong feelings about probability. Coincidentally, it's something Tom mentioned in his column, and it intersects with the questions I asked myself about the SHSS risk level. I couldn't immediately think of an example of a SHSS accident, and the email responses quickly gave me several examples. After reading these, I recalled the AC-130J, a recent example I should not have forgotten. The reminder affected my perception of my own objectivity. How do we assess the probability of a given hazard? More importantly, though, how confident do we feel about our probability assessment?

Perhaps both of these questions should have been asked by the C-17 test team introduced in the last issue. As you may recall, that issue introduced the topic, Communicating Uncertainty in Flight Test, by revisiting Safety Planning in C-17 Airdrop Flight Test: The test pilots used a C-17 HITL (hardware-in-the-loop) simulator to prepare for an airdrop envelope expansion and handling qualities test. The simulator did not contain models for the test items (i.e., cargo and airdrop parachutes), raising questions about the results. In the example test campaign, the test team determined upper and lower bounds on the response of the aircraft during contingency situations. This is different from the normal procedure of predicting the estimated response of the aircraft. This technique may not work in every situation, but it illustrates the use of heuristics, which, as many have suggested recently, ought to play an important role in helping the flight test profession cope with complexity and uncertainty.

This past example and the SHSS risk problem frame the critical question: **How do we express confidence in model, simulation, and experimental outcomes?** To answer this question, I propose a framework of three rules, specifically for communicating about uncertainty in flight test outcomes. It may be helpful to abbreviate these heuristics as the **3Q**.

The purpose of this article is to introduce the rules and demonstrate their utility and application:

1. Express the outcome both qualitatively and quantitatively.
2. Describe the range of possible outcomes.
3. Assess the frequency of potential outcomes.

To illustrate these concepts, I want us to recall some common characteristics of an airplane flight manual. Most of the people in this audience have heard of Notes, Cautions, and Warnings,

especially in this context. One such manual defines these terms in the following way.
WARNING denotes those items highlighted for the purpose of describing "operating procedures or techniques which may result in personal injury or loss of life if not carefully followed."
CAUTION denotes those items highlighted for the purpose of describing "operating procedures or techniques which may result in damage to equipment if not carefully followed."
NOTE is "additional or significant operating information requiring emphasis."
These definitions do not use numerical descriptors, yet after reading them, we come away with an understanding of their relationship, relative severity, and importance. Instead of being quantitative, the terms are qualitative.

There are many more examples of the meaning of qualitative: yes or no, high or low, left or right, and even first or last. Qualitative does not necessarily mean that something is indefinite or imprecise, but instead, it is the complement of—it contrasts with—the idea of quantitative, or numerical and measurable characteristics, a term which needs no illustration.

This establishes precisely what we mean by qualitative, which is our foundation in the subject of communicating and understanding uncertainty. To help us build on this foundation and create a shared lexicon for dealing with the varied topics under the broad umbrella of uncertainty, I would like to propose three foundational rules. First, we should express ideas both qualitatively, as we did above, and quantitatively. Second, we should attempt to describe the range of possible outcomes. Finally, we ought to assess the frequency of potential outcomes.

The figure above places the qualitative phrases from the flight manual on a spectrum of possible outcomes, in relation to one another. This illustrates the second rule, describing the range of possible outcomes. This spectrum can also help us understand the third rule. We do not expect that "loss of life" will happen frequently in the service life of an airplane, but the conditions described in a note might happen daily. This is what we mean by "frequency of potential outcomes."

Having briefly explained each of the rules, now apply these two ideas to safety process outcomes. In a Test Hazard Analysis, we immediately see that the characterization of hazards and their severity aligns naturally with the qualitative. Death or injury or aircraft damage are three terms used often that describe the severity of outcomes of a particular hazard. Furthermore, the formal FAA and ICAO terms used to describe hazard severity are qualitative (catastrophic, major, minor, etc.). This much is not new to most readers, but let us explore the idea further.

Is it possible to define hazard severity quantitatively rather than qualitatively? One may struggle initially with such a definition (*I certainly did). Consider, however, this simple characterization: A safety outcome or test hazard that results in zero injuries or death is not severe. A severe outcome would have some positive number of injuries or death. We could further delineate by calling a hazard very severe if death (not just injury) is an outcome. This gives us three levels of severity defined quantitatively—not severe, severe, and very severe. We simply extended our qualitative ideas (injury or death) and assigned numeric levels to quantify them. Aircraft damage also falls into this framework. Quantitative characterization extends even further, allowing us to assign continuous, even infinite, values, as well as precision and resolution to our understanding of safety process outcomes.

We can also apply this same contrast to the causes of a hazard. For example, consider a hazard such as aircraft damage, the cause of which is overspeed. We have defined this cause qualitatively as "overspeed," but we can extend the definition quantitatively. A stable test point conducted at VMO that encounters a light atmospheric disturbance may result in a momentary overspeed of 1-2 knots that immediately returns to its steady-state value. But this is probably not what we imagine when we consider the cause of such a hazard. To create a more precise definition, we could further quantify our meaning.

Conclusion: Last year, Doug "Beaker" Wickert declared that we needed more time to think—I agree and would further suggest that it is the intentional cognitive exercise presented herein from which we benefit. By pondering these 3Q, we exercise our thinking. Precision and accuracy of definition are important, and the exercise forces us to transform vague notions into specific definitions with qualitative characteristics. Finally, the stories told by Rod and the FTSC and presented at workshops give us data we can directly use to assess frequency and range of potential outcomes. *Mark Jones Jr.*