Flight Test *Safety* Fact

Published for the Flight Test Safety Committee

Chairman's Comments: NTSB Reports on Uber Safety Culture *Tom Huff*

Greetings fellow active and former flight testers, aspiring test professionals, safety practitioners and test organization leadership. As we pen this December edition of Flight Test Safety Fact (FTSF), we hold ourselves to account on how we're doing in terms of outreach. Thank you to all who help us make this a persistent effort. Reach the unreachable; reach everyone.

In preparing for each FTSF, Mark and I typically bounce ideas of each other to make sure the content is relevant and sparks introspection, discussion, and debate. The Flight Test Safety Committee Board members weigh-in as well. This sausage-making stems from a culture of open dialogue, ultimately, our way of formulating strategies to make our industry safer. In my day to day, I consume quite a bit of aviation safety media reporting and frequently use that as a basis for my brainstorming. Thus, it was a recent media release from the NTSB that caught my attention: "Inadequate Safety Culture Contributed to Uber Automated Test Vehicle Crash" (full text of press release included below.)

Those that know me, know I'm strongly opinionated about the effects of *culture* on an organization. Furthermore, I have gone on record to suggest that without a positive safety culture, your Safety Management System (SMS) will be a paper tiger. You simply will not reap the safety performance benefits without first establishing a strong cultural foundation. The cultural rebar - a critical component of the foundation to assure strength - has to come from leadership: FROM - THE - TOP. It must then transcend to every level of the organization. When I discuss SMS auditing with different organizations, I usually ask, "Does your CEO get interviewed?" The reply "Say what?" is the usual response. "Oh no, s/he is far too busy." That comment alone tells me that the production - protection balance could be out of whack. Fate is lurking right around the corner...

Ultimately, the NTSB found that the [ground] vehicle automated test resulted in risk realization (catastrophic) with an innocent non-participant [needlessly] losing their life. The primary 'mitigation' appears to be distracted at the wheel.

Obviously, the NTSB was not amused at the level of risk awareness and accident prevention strategies. Disabling native safety features should have been a huge red flag for safety reviewers and risk accepters. Perhaps these were viewed as "acceptable" given the presence of a human operator. How many accidents have we seen when we "assume" that the human will intervene in the expected way?

(continued page 3)

In This Issue

Artificial Intelligence – Introduction to AI and flight test safety
Reach Everyone – How much progress have we made in 2019? How many is "117%"? Will you help us Reach Everyone?
Chairman's Comments – NTSB report on Uber safety culture
Contact FTSC – We've added AIAA contact info! Derek will slowly take over for Starr Ginn, who chaired this past year.

Contact 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
AIAA Flight Test Group derek.spear@gmail.com

Contact *Flight Test Safety Fact*
Mark Jones Jr, Editor mark@flighttestfact.com

Reach Everyone *Mark Jones Jr.*

More than a year ago, I heard a long-time USAF member of SETP say, "I never heard about the Flight Test Safety Committee" (an admission that took many of our Committee members by surprise). That conversation was the catalyst for this newsletter, which distributed its first issue in January this year. On the other hand, since January, someone confessed that they had NEVER heard about the Flight Test Safety Committee, even though they attended more than 15 SETP Symposia in Anaheim and worked with SETP, SFTE, and AIAA members for more than 20 years. I'm actually delighted by the second story because they found out about the Committee through this newsletter--our efforts over the past year have succeeded: We've made progress!

Our explicit goals is to Reach Everyone with this publication. This is a qualitative statement that exaggerates the importance of getting the word out. More specifically though, we intended to reach 117% of the membership of SFTE and SETP in the first year. So here we are, and it's time to ask, "How did we do?"

We reached 3539 people by direct email with our newsletter, as seen in the table below, which shows "total membership" data—the FTSC consists of AIAA, SFTE, and SETP members.

The first column shows current total membership as of November 18, 2019. The second column shows the highest membership reported over the course of the year, and the final column reports the total number of email addresses on file. Not all members provide this contact info.

(continued next page)

Reach Everyone: 117% (continued)

The final row, "Other," are those who receive the newsletter because they asked us to add them to the distribution or attended the Flight Test Safety Workshop. The data shows that our email distribution has grownslightly more than over 3%, to 103% total, which brings us short of our goal. This number does not include people who receive the email forwarded by an existing member.

2019 FTSC Membership Data			
	Current[1]	**Highest[2]**	**Email[3]**
FTSC	**3597**	**3614**	**3429**
Other	**110**	**110**	**110**

Notes:
1. Membership data current as of November 18, 2019.
2. Highest reported membership year-to-date.
3. Not all members provide email addresses.

In the rest of this article, I want to accomplish two things. The first thing is to introduce a technical topic relevant to estimating progress toward our 117% goal, a topic we will later explore for its relevance to estimating risk in flight test safety. Secondly, I want to finish evaluating progress toward our goal and communicate a vector for next year. You can now skip to either column below, left or right, and continue reading, or you can read both columns.

Reach Everyone (117%)

Implicit in the statement of this goal is the need to count how many members are in SFTE and SETP. Sometimes, however, the simple words used to make a statement hide the technical challenges of accomplishing a task like this.

Did we mean the number of members in January or in December, or did we mean the highest membership number reported for the year (that sounds reasonable)? Are we counting members of both Societies twice? It is a simple but important question, just the first step in actually measuring the quantity.

Another, more recent example is also relevant. One report from the 2019 SETP Symposium stated 692 people were in attendance, but a spreadsheet of the registrants explicitly listed only 674 people. Each of us could probably explain several possible causes for the discrepancy, but this example illustrates something we actually overlook: "Counting is hard."

Estimating Risk in Flight Test Safety

The recent pad abort test of Boeing's CST 100 Starliner provides an example we can use to discuss the estimation of risk. During the flight test, only two of the three recovery parachutes deployed successfully.

Thinking about these kinds of failures is exactly what we do when performing test hazard analyses. We say things like "unlikely to occur" when we are talking about the probability—or frequency or likelihood—of hazards. Implicit in that language is the necessity of counting how many times a given failure occurs in a given number of flight hours.

In the Starliner case, do we count this test as a failure, and if so, how many failures should we expect in the lifetime of the capsule? Since the test team quickly ascertained root cause, does it change the total or how we count failures? This example illustrates something we actually overlook: "Counting is hard."

Maybe you finished one or both of those columns and read the conclusion with incredulity. You may wonder where this is going. Ultimately, I want to be able to measure how many people we reached with the news: News about the Flight Test Safety Committee (FTSC), its members and their presentations, resources, workshops, and the newsletter; news about new techniques, training, and trends. We want to be able to spread news that will have a direct impact and will equip flight test professionals with the knowledge and skill to make life-saving decisions. That's where this is going. That requires us to *count* how many people read the newsletter, whether they subscribe to it or a colleague forwarded it. This turns out to be very similar, statistically speaking, to estimating the number of failures that may affect a flight test—counting failures is a topic we will address more completely in a future issue.

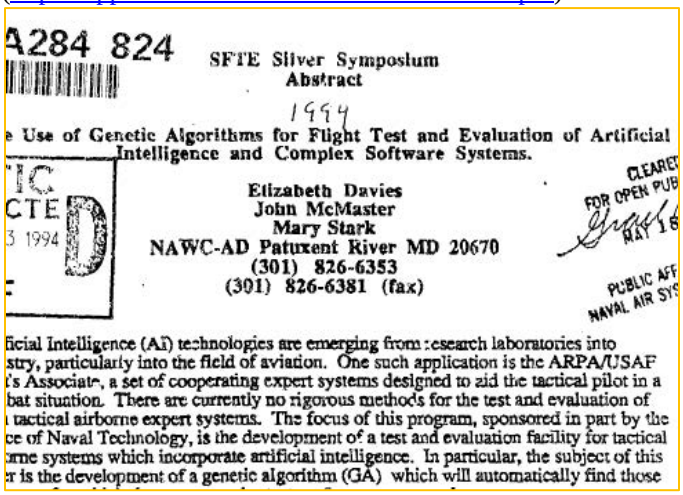
You may recall that one of the "3Q" heuristic rules recommended for evaluating uncertainty is express the outcome both qualitatively and quantitatively. So I'll begin with a qualitative report of newsletter reach during 2019. I began by polling the FTSC: There are 21, and several of them have forwarded the email to their entire office: 25 people in one and 31 in another, for example. How many of those people are "unique," i.e., not included in FTSC membership numbers? Unfortunately, we don't know, but some statistical counting techniques can address this uncertainty.

For example, we can "estimate the range of possible outcomes." One way to do this is count each of the reports above as unique "hits." This gives us an *upper bound* on how many we've reached—remember we are merely estimating the top of the range. This results in 61 more, another 2%. Another variation would be to ascribe a frequency of more than 1 to some of the unique hits above: For example, someone from RAAF Aircraft Research and Development Unit (ARDU) subscribed and we could count this as two hits (him and someone in his office). This is a reasonable way to estimate the reach of this particular subscriber.

Each of these gives incremental progress, but we need to reach 5x as many new readers to attain our goal of 117%. Next year, we plan to incorporate distribution to the AIAA Flight Test group and count their numbers as well. The target will continue to be 117% for the second year as we refine our data collection. Next year, we will cover counting techniques in more detail and show their relation to "counting failures" in flight test. In closing, though, let me reiterate this: For each example above, an existing members took the initiative to share this newsletter. **Thank you! Your effort will help us continue to make progress toward our goal.**

Artificial Intelligence: A Limited Survey

As far back as 1994, Flight Test Safety intersected with Artificial Intelligence (AI). As you can see below, that year the US Navy at Pax River published a paper titled "The Use of Genetic Algorithms for Flight Test and Evaluation of Artificial Intelligence and Complex Software Systems" (https://apps.dtic.mil/dtic/tr/fulltext/u2/a284824.pdf).



A284 824

SFTE Sliver Symposium
Abstract

1994

e Use of Genetic Algorithms for Flight Test and Evaluation of Artificial Intelligence and Complex Software Systems.

Elizabeth Davies
John McMaster
Mary Stark
NAWC-AD Patuxent River MD 20670
(301) 826-6353
(301) 826-6381 (fax)

CLEARED FOR OPEN PUB
MAY 18

PUBLIC AFF
NAVAL AIR SYS

ficial Intelligence (AI) technologies are emerging from research laboratories into stry, particularly into the field of aviation. One such application is the ARPA/USAF 's Associate, a set of cooperating expert systems designed to aid the tactical pilot in a bat situation. There are currently no rigorous methods for the test and evaluation of tactical airborne expert systems. The focus of this program, sponsored in part by the ce of Naval Technology, is the development of a test and evaluation facility for tactical rne systems which incorporate artificial intelligence. In particular, the subject of this r is the development of a genetic algorithm (GA) which will automatically find those

A lot has changed since then. One of the most informative sources (including technical details) of ongoing progress in this field is the Airbus Vahana team. They performed the first flight of the Vahana aircraft on 31 January 2018. In May of this year, they completed full transition from hover to forward flight.
Here are links to a series of articles they've published.
Vahana First Flight - https://vahana.aero/vahanas-first-flight-a-success-ade26d26ba02.
May 2019: Full Transition - https://vahana.aero/flight-test-update-full-transition-5e2f686c22b4.
The final story in their series: https://vahana.aero/our-story-part-4-7d8cec453408.

Vahana First Flight (Airbus photo)

This year, SFTE's Annual Symposium featured two other papers on very closely related topics.

Machine Learning Techniques Applied to Flight Test Data Evaluation, by Kelton Busby (kbusby@aerotec.com) and Rebecca Hattery (rhattery@aerotec.com).

Framework for Safe, Effective, and Efficient Testing of Autonomy, by Captains Riley Livermore and Richard Agbeyibor, Emerging Technologies CTF (Combined Test Force), 412th Test Wing, Edwards AFB, CA.
At press time, these authors were not available to share these papers herein.

The final two headlines comes from Thales, both of which jumped out at me as I prepared this limited survey.

(Photo credit: Thales.)

Psibernetix is an Ohio-based company that writes software (using AI) to perform V&V (Verification and Validation) on AI software. Thales recently acquired them as described here: https://www.aviationtoday.com/2019/11/27/thales-expands-psibernetix-ai-smarter-fighter-pilot-training-software/.
Thales has developed a FMS from scratch that incorporates AI: https://www.ainonline.com/aviation-news/aerospace/2019-11-26/thales-reveals-next-generation-fms-pureflyt or http://onboard.thalesgroup.com/now-on-air-introducing-pureflyt-the-new-generation-flight-management-system/.

Chairman's Comments (continued)

Reference the NTSB recommendations, the Congressional Research Services Report and Joint Authorities Technical Review (JATR) on the 737MAX for in-depth discussion on this. And yes, I do think it important to consider human performance in both the assumptions made in the conduct of test as well as during in-service operation of the product.

Returning to culture: This accident should give us pause because it's a test accident, and we've learned similar lessons where culture was a negative influence. With the proliferation of unmanned vehicles and the desire to rush-to-market, schedule pressure seems to compromise rigorous risk assessing and measures to ensure risk is reduced as low as reasonably practicable (ALARP). Yes, I chose words carefully there, using the common nomenclature—ALARP—to ask: is this is too ill-defined? How rigorous is the attempt to achieve ALARP, and then who decides when you are there? Perhaps a topic for another time... *(continued next page)*

SFTE Holds Final Safety Webinar

The SFTE Tech Council hosted the last Flight Test Safety webinar and presented it to SFTE members throughout the country this last week in November. This four-part webinar series started in March of this year as an idea conceived by SFTE Fellow Dave Gibbins, who recognized that few of the current FTEs got the opportunities to experience things that his generation did and suggested this venue as a way to share this information. Bob Barham supported the effort with Al Lawless, chair of the Tech Council. If you missed the live webcast, you can still access the slides and video. To view a recording of the Flight Test Safety Red Flags Webinar, go to the SFTE website: http://www.sfte.org/tech-council/. Pictured here is a listing of available material. This is an excellent source for safety continuation training or professional development.

A total of seven flight test accidents were analyzed and discussed to determine "common threads," so lessons learned could be obtained, in the hope that they do not repeat in the future. The following table indicates the aircraft of the flight test accident and the author of the summary. The authors did an excellent job summarizing the accident in a clear way and then researched the accident reports to summarize the findings.

Flight Test Safety Case Study

Aircraft	Author
X-31	Al Lawless
A340-600	Jim Fawcett
YF-22A	Bob Barham
Challenger CL-600	Jim Martin
B-1A	Otto Waniczek
G-650	Al Lawless
Ranger 2000	Al Lawless

The table below summarizes the contributing factors in that accident sorted by airplane model. You can see on some accidents there were several issues. *(continued next page)*

Chairman's Comments **(continued)**

In conclusion please help us improve and further the awareness of flight test resources that can make us safer. Share the newsletter, leverage the Societies web content, and continue the dialogue on challenging issues. I might suggest that an organization with strong safety culture might already be sharing the newsletter. If you are a fan of SMS—and I hope you are and have one—this would be part of your Safety Promotion efforts. You do have a Learning Culture right?

We welcome input and differing points of view. Don't be shy about sharing your thoughts, and anonymity will be extended if requested. Send us your thoughts at: chairman@flighttestsafety.org.

In your service,

Tom Huff, Chairman, Flight Test Safety Committee

"Inadequate Safety Culture" Contributed to Uber Automated Test Vehicle Crash - NTSB Calls for Federal Review Process for Automated Vehicle Testing on Public Roads (NTSB Press Release, 11/19/2019)

https://ntsb.gov/news/press-releases/Pages/NR20191119c.aspx

WASHINGTON (Nov. 19, 2019) - The National Transportation Safety Board called upon federal regulators Tuesday to create a review process before allowing automated test vehicles to operate on public roads, based upon the agency's investigation of a fatal collision between an Uber automated test vehicle and a pedestrian. During a board meeting held to determine the probable cause of the March 18, 2018, Tempe, Arizona crash, the NTSB said an Uber Technologies Inc. division's "inadequate safety culture" contributed to the March 18, 2018, nighttime fatal collision between an Uber automated test vehicle and a pedestrian. The vehicle operator was uninjured in the crash; the pedestrian died. Uber's Advanced Technologies Group had modified the striking vehicle, a 2017 Volvo XC90, with a proprietary developmental automated driving system. The vehicle's factory-installed forward collision warning and automatic emergency braking systems were deactivated during the operation of the automated system. An Uber ATG operator was in the driver's seat, but the automated system was controlling the vehicle when it struck the pedestrian at 39 mph. The NTSB determined that the immediate cause of the collision was the failure of the Uber ATG operator to closely monitor the road and the operation of the automated driving system because the operator was visually distracted throughout the trip by a personal cell phone. Contributing to the crash was Uber ATG's inadequate safety risk assessment procedures, ineffective oversight of the vehicle operators and a lack of adequate mechanisms for addressing operators' automation complacency – all consequences of the division's inadequate safety culture.

"Safety starts at the top," said NTSB Chairman Robert L. Sumwalt. "The collision was the last link of a long chain of actions and decisions made by an organization that unfortunately did not make safety the top priority.

Recurring Flight Test Safety "Red Flags" Summary

Red Flag	X-31	YF-22	B-1A	Silver Bird	A340	CL-600	G650	Ranger 2000
Complacency	✓	✓						
Abnormal Ops	✓		✓					
Configuration Control Errors	✓							
Novel, Amateur Built				✓				
Ineffective Communications		✓						
Time and Commercial Pressure					✓	✓	✓	
Inadequate Procedural Doc					✓			
Mis-matched crewing			✓					
Warning Fatigue			✓					
Crew Preoccupation			✓			✓		✓
Marginal Performance in High Risk Testing						✓	✓	✓
Safety Critical FT Eqpt						✓		
Insufficient Tech Knowledge	✓				✓		✓	
Incorrect Safety Assumptions							✓	
Upper management disassociation							✓	

How to Use

Take a look and then consider hosting a session to share one webinar with your teams. Afterward, develop mitigations for each of the "Red Flags" for your organization to use and to share. A new employee, intern, and even current college students could use these as well: do their own research under the supervisor's guidance and a fresh perspective, a new set of eyes *Dan Hrehov*

"Counting is Hard" - Reach Everyone Summary

3Q Rule for Communicating Uncertainty	"Reach Everyone" Outcome
Express the outcome both qualitatively and quantitatively.	Newsletter reached 110 more people by direct email than we reached by direct email in SETP and SFTE combined. This is approximately 103% of SETP and SFTE combined membership, less than our goal of 117%. We reached 3539 by direct email, less than the combined membership total of 3614. We reached new people and new organizations almost every month. Some members achieved almost 300% sharing reach.
Describe the range of possible outcomes.	Some recipients of the email are of members of both Societies. SETP did a study estimating the number of test pilots from Turkey to the Pacific Rim, and the data from this study suggests there are 460-690 additional flight test professionals outside of our current reach.
Assess the frequency of potential outcomes.	When we reach a "new organization", we can reasonably assess a frequency of 2 (one for the Chief of Safety, for example, and one for another person in his office)—this gives us a reasonable lower bound on frequency. Another way to assess the frequency is to use "average office size" based on reported data from FTSC members. A third variation would be to assess a given number for each new organization based on a reasonable estimate of personnel in that organization.