

The Test Like You Fly and Test What You Fly Approach for the Artemis Human Spaceflight Paradigm

Tracy R. Gill , John S. Gurecki , Dr. Eirik Holbert , Syrus K Jeanes , Robert M. Peacock , and Joseph M. Schuh, Ben Jimenea, James R. Debruin, Jody H. Fluhr (employed by BAH), Christy Gattis

NASA

AIAA SciTech, January 12-16, 2026

Published by the American Institute of Aeronautics and Astronautics, Inc., with permission.

Test Like You Fly (TLYF) Introduction

“

The aerospace industry is unique... because a thousandths-of-an-inch mistake can cause spectacular failures.”

– Mae Jemison, astronaut.

“If you can't afford to do something right, then be darn sure you can afford to do it wrong.”

– Industry aphorism

“You must build a coalition of the willing to bring forth cultural change for Test Like You Fly to ensure crew safety and mission success.”

– Tracy Gill

Test Like You Fly (TLYF)

The principles of Test Like You Fly (TLYF) and Test What You Fly (TWYF) seem self-explanatory—the more closely preflight testing can replicate flight conditions and mission operations, the less risk of unexpected issues during execution of the mission.

However, for every spaceflight program, a multitude of factors dictate the extent to which that principle can be followed. Therefore, the TLYF process involves a continual assessment to maximize the effectiveness of preflight testing, given inevitable limitations, and to characterize the remaining risk to crew safety and mission success.



TLYF Process: Why Now?

NASA has performed test like you fly assessments throughout every human-rated and science program/mission to manage risk. We have learned hard lessons and committed to advancing in honor of those missions and crew we have lost. Going back to the moon is hard. Going on to Mars will be even harder.

Although we have performed TLYF assessments and documented exceptions for every program/mission, a methodical approach for the assessment or acceptance of the risks associated with the exceptions would help drive out more specific risks and allow for optimization. The first step in optimizing the process is to write it down.

Without ensuring a comprehensive evaluation of the system(s) and a structured consideration of the risks, we cannot truly understand how TLYF shortfalls may have contributed to issues or failures and fully understand risks associated with future TLYF exceptions. Not only can we strive for consistency across our agency's enterprises, but we can move to improve upon it.

We must ensure we are asking the right questions to ensure crew safety and mission success.

As our workforce ages, many will leave and take decades of experience with them. Coupled with increased pressure on discretionary spending and a mindset of continually doing more with less, standardization and optimization of crucial processes is more important than ever.

NASA's Moon to Mars Architecture

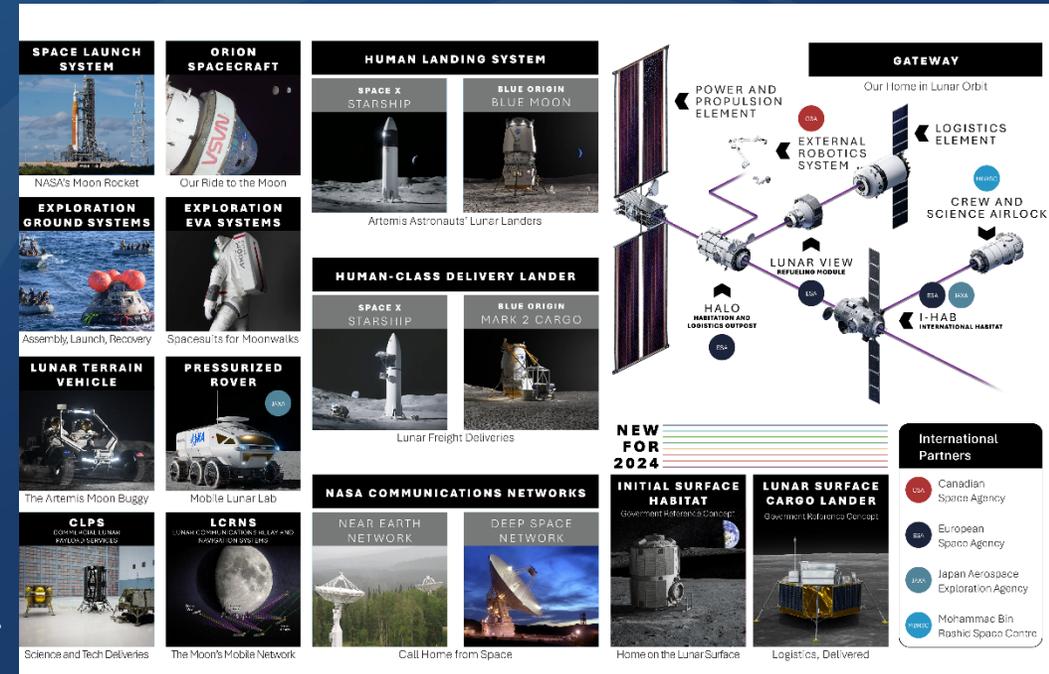


Image Credit: NASA

Case Studies

Examples of what has occurred in the past that drove TLYF implementation

Case Studies of Previous TLYF Outcomes

- NASA has been “testing like you fly” throughout many of the human spaceflight programs, for design certification as well as operational and mission readiness scenarios.
- A variety of challenges drove these preflight ground tests, such as new technologies, new types of interfaces, and previous successes or failures.
- While NASA did not develop a consolidated TLYF lessons learned product, there are papers and mishap reports referenced herein that discuss using TLYF to increase the likelihood of mission success and crew safety.
- Test Like You Fly finds its genesis from a need for crew safety and mission success.

Space Shuttle Main Engine Design Certification Testing

- The Space Shuttle Main Engine (SSME) Project used ground testing to support both certification and acceptance testing. SSME testing in support of design certification included testing of prototype engine and several development engines, including 150 engine firings with a cumulative run time of 3,500 seconds completed.
- The tests uncovered significant technical issues, especially with the engine's high-pressure fuel turbopump. Testing continued with development engine firings evolving to full 100 percent rated power for a full flight duration of 520 seconds. By the end of the development testing phase, the SSME test program had accumulated more than 34,000 seconds in 394 firings. [4]
- Single engine tests of SSMEs continued through the life of the Space Shuttle Program in support of build verification as well as block upgrades associated with six unique configurations.
- This ongoing campaign continually enhanced reliability and safety of the SSMEs and reduced overall mission and crew safety risk. The block upgrades were driven by ground test failures, in-flight anomalies, and safety assessments which identified critical failure modes.
- Until the final SSME test in 2009, 2,307 test firings were completed, representing 228 hours of hot fire time. This campaign represented a significant investment of time and resources, but the outcome was a realized reliability of 1 for the SSMEs during Space Shuttle missions. [4]

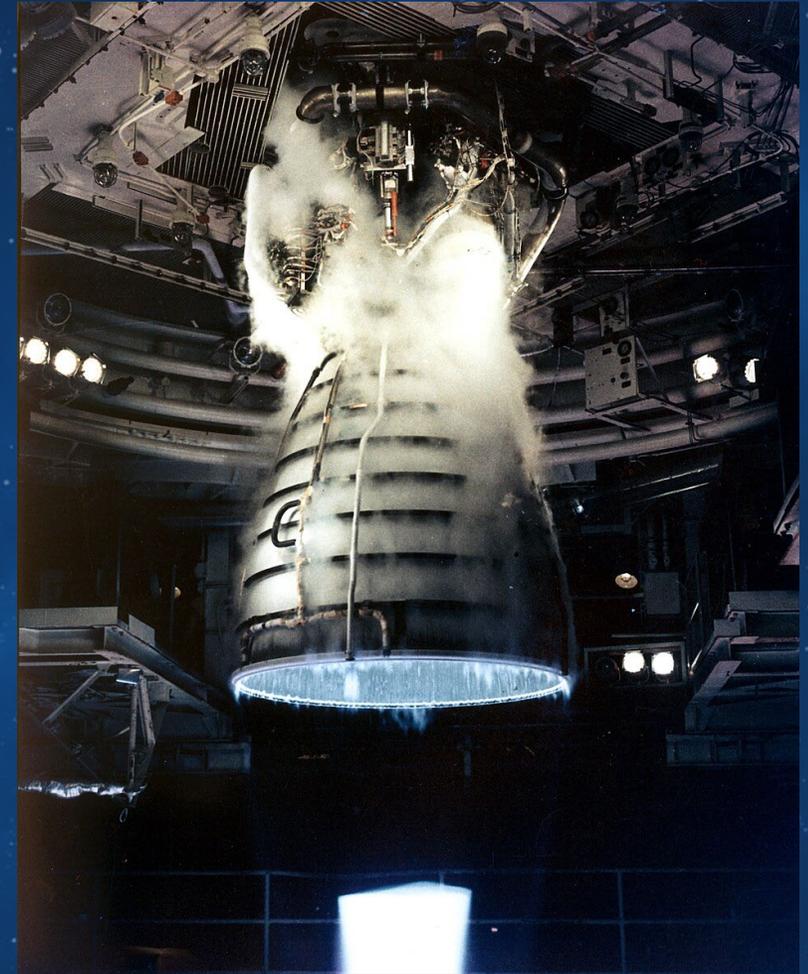


Image Credit: NASA

Main Propulsion Test Article

- While single-engine tests remained critical in the SSME testing and certification program, multi-engine testing prior to the first shuttle flight provided a more “like you fly” scenario of the entire propulsion system would behavior during an actual launch.
- The Main Propulsion Test Article (MPTA) -098 included a flight-weight shuttle aft fuselage accommodating three SSMEs, a simulated orbiter midbody, and an attached flight-like External Tank (ET) to deliver liquid hydrogen and liquid oxygen to the engines. [4]
- MPTA testing prior to STS-1 identified at least 4 significant failures. The resolution included redesigning several catastrophic failure modes out of the system design, enhanced inspection of brazed joints associated with the nozzle, increased control/testing of rod weld material, and enhanced inspection of liquid oxygen post concentricity. [5] These flaws, if not identified during a very flight-like test scenario, could have resulted in loss of the initial Shuttle mission, vehicle, and/or crew.
- In total, in preparation for STS-1, engineers tested 21 engines, including the three that flew on Columbia’s first flight, over the course of 575 single-engine tests and the 18 MPTA tests, accruing more than 87,000 seconds of engine run time. As noted in examples above, these efforts contributed to a very successful mission life for SSME. [4]

Main Propulsion Test Article being lifted onto its test stand in 1977



[Image Credit: NASA](#)

Space Shuttle Orbiter Flight Readiness Firings

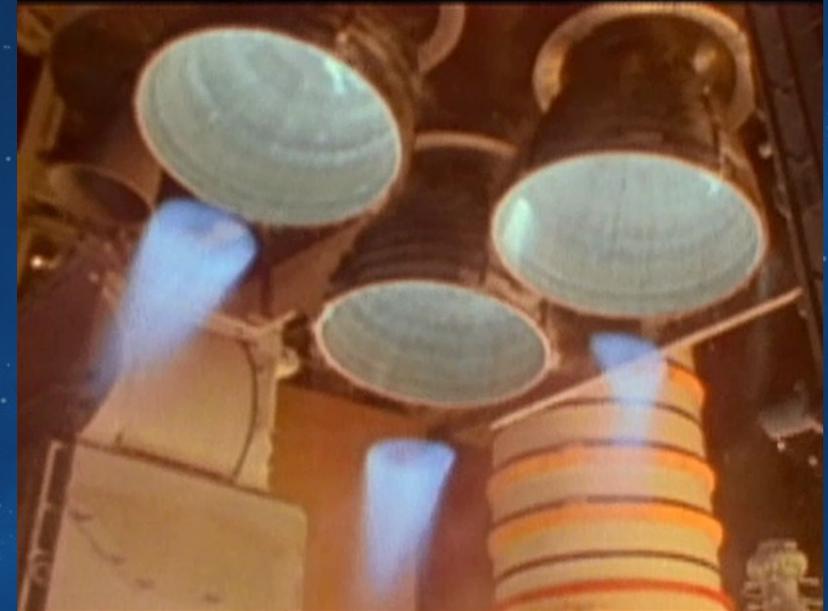
- Test Like You Fly testing is important for flight hardware and software/firmware that will fly in support of a mission. This testing is sometimes called “acceptance” or “build verification.”
- One very important test that was performed at least once on each newly built Space Shuttle Orbiter was the Flight Readiness Firing (FRF).
- Flight Readiness Firings (FRFs) were performed on seven occasions during the shuttle program’s 30-year history and sought to validate all the vehicle’s systems under the closest possible conditions to actual flight, without actually leaving Earth.
- Cost and schedule implications are often drivers when planning—or not planning—large-scale integrated system tests.
- History has shown us that these tests are very useful in finding off-nominal system level interactions which would not have been discovered, and been able to be mitigated, in lower-level tests.



NASA Photo KSC-81PC-101

Shuttle Flight STS-93 Anomalies

- During STS-93, there were two serious in-flight anomalies, resulting in NASA leadership chartering the Shuttle Independent Assessment Team (SIAT). The SIAT's charter was to bring to Shuttle maintenance and operations processes a perspective from the best practices of the external aviation community, and report to the Associate Administrator (Office of Space Flight) [6].
- Issue #5 in the report was titled “The SSP should adhere to a “fly what you test/test what you fly” methodology.” In the Propulsion section of the report, it states: “The SIAT considers that a serious lapse in judgment and/or in attention to the engine data base occurred, which allowed two pins to be used in STS-93, without ground test verification firing.” Columbia's right engine had two of these pins jammed into oxygen-injection tubes as plugs to keep liquid oxygen from entering the damaged lines.
- Handling of the pin insertion and test as a standard repair, which did not necessitate a hot fire retest, precluded management visibility of the frequency of liquid oxygen (LOX) post deactivation.
- Of 19 pins ejected during ground testing, all but one were ejected during the first engine firing. This process of repair without retest was determined to have significantly contributed to the subsequent pin ejection and the nozzle damage during STS-93 flight. Fortunately, this deviation from TLYF the STS-93 only resulted in an abort to orbit versus a more serious condition [6].



[Image Credit: NASA/Johnson Space Center](#)

International Space Station Express Logistics Carrier

- Prior to first ELC launch, STS-129, a cost reduction decision was made not to loan the ground unit of the automated payload switch (APS) from a JSC lab to KSC for ground processing.
- Instead, a patch panel was used for testing. Concurrently, an engineering change order incorrectly swapped High Rate Data Line (HRDL) Rx and Tx lines on ELCs. • Thus, ELCs were tested in a different configuration than the AMS-02 payload with the same interfaces.
- In-orbit, it was discovered that fiber optic HRDL links for ELC 1 and 2 had Rx and Tx swapped. A retrofit had to be designed, built, launched, tested, and installed by astronauts during IVA.
- Cost and time of total retrofit, and repair effort would exceed cost and time of initial shipment and test request by orders of magnitude.
- Relatively speaking, this was readily correctable as the failure occurred on the ISS. If this happened in the vicinity of the Moon, the cost, effort and consequences could be far worse and potentially unrecoverable on that mission.

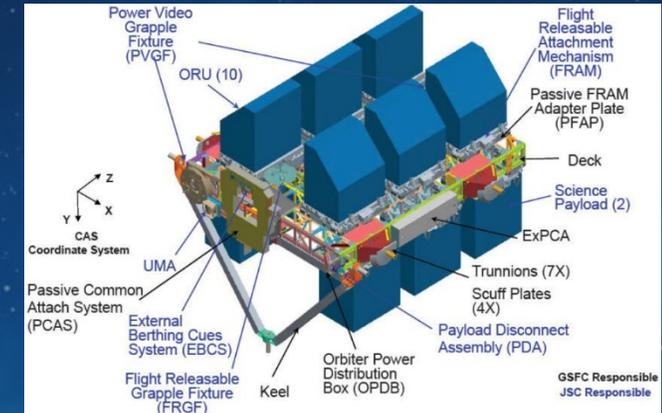
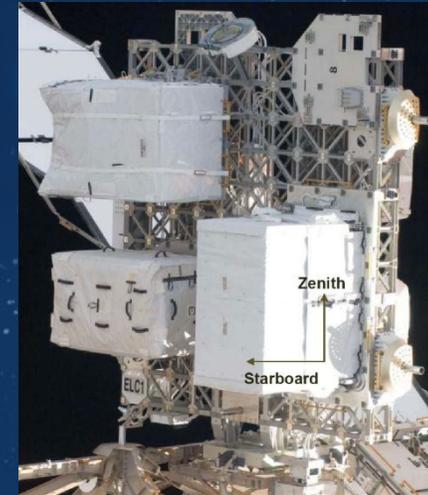


Image Credits: NASA

ISS Multi-Element Integrated Testing (MEIT)

- The initial US Lab element critical activation took 36 hours, and the requirement was less than two hours. The impact could have resulted in the loss of the US Lab element during on-orbit activation due to thermal loading timeline requirements. Loss of Mission (LOM)/Loss of Vehicle (LOV) – no crew safety concerns.
- An electrical short was identified/corrected on the Flight Support Equipment Grapple Fixture (FSEGF). The impact could have resulted in failure of the Space Station Remote Manipulator System (SSRMS) to power-up. Minimal impact would have been an unscheduled EVA to identify and correct the problem. Maximum impact would have been a SSRMS shuttle re-flight (LOM).
- The Assembly Power Converter Unit (APCU) was redesigned after power losses due to breaker trips during the ISS P6 element activation on the ground before flight. The impact would have prevented the activation of P6 on-orbit, driving a shuttle re-flight of P6 (LOM).
- Video lines were identified as swapped between Trailing Umbilical Systems 1, 2 (US) and Mobile Base System (Canada). Significant operational impact to manually route video signals from Space Station Robotic Manipulator System (SSRMS) to Remote Work Station (RWS). EVA would have been required to replace the harnesses.
- Metrics show that later testing such as Node 2 IST and MEIT3 revealed significant improvements in the software products as well as the hardware delivered from the factory. For MEIT1, a total of 884 Interim Problem Reports (IPRs) were taken, while MEIT2 had 334, and MEIT3 had 64. As the test program matured, the earlier verification testing, software products, etc. that fed into the MEIT campaigns improved resulting in fewer difficulties during the integrated testing

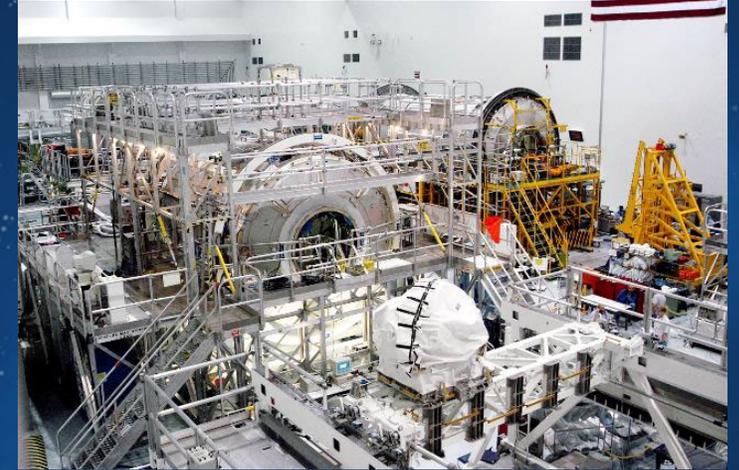


Image Credits: NASA

Moving from Lessons Learned to Lessons *Implemented* with TLYF Process

"If you can't describe what you are doing as a process, you don't know what you're doing."

-W. Edwards Deming

"We are what we repeatedly do. Excellence, then, is not an act, but a habit."

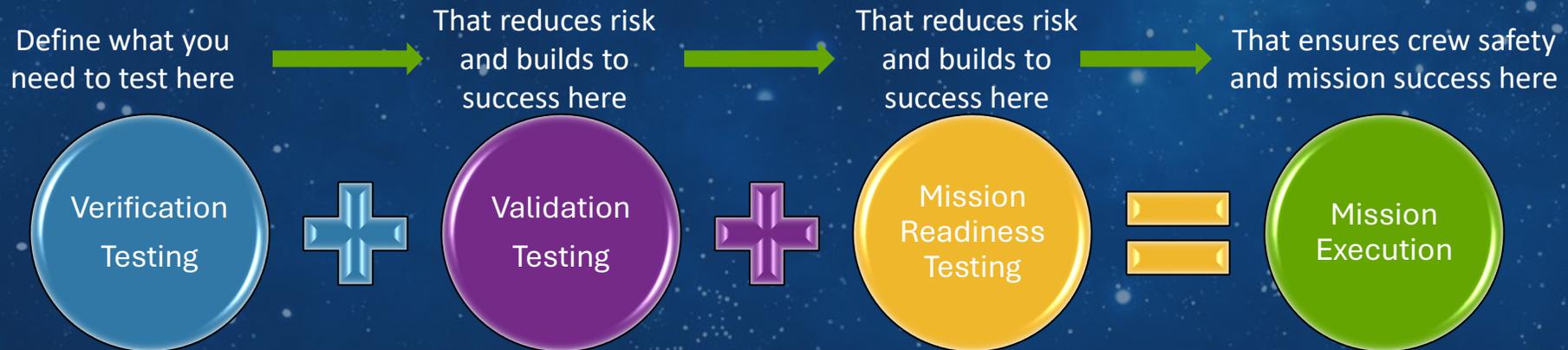
- Aristotle

TLYF Process Goals

The first area focuses on enabling verification test opportunities, including establishing a test configuration along with the means and assets required to execute that test configuration.

The second area is focused on identifying the high priority test operations to perform for validation of interfaces, operability, and scenarios that address TLYF, Test Like You Operate, and mission readiness perspectives.

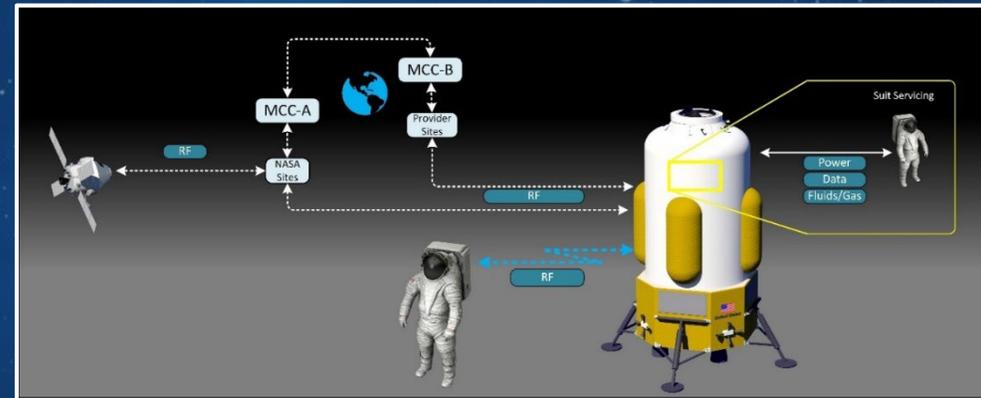
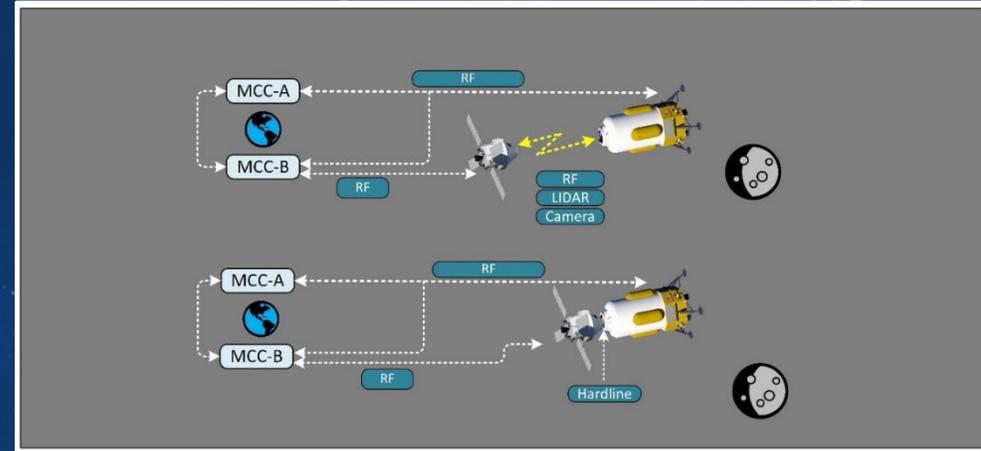
Verification and Validation testing build up to Mission Readiness testing which share the common goal of reducing risk to the crew and mission.



← What is the difference between the test activities on the left and the flight activities on the right? →

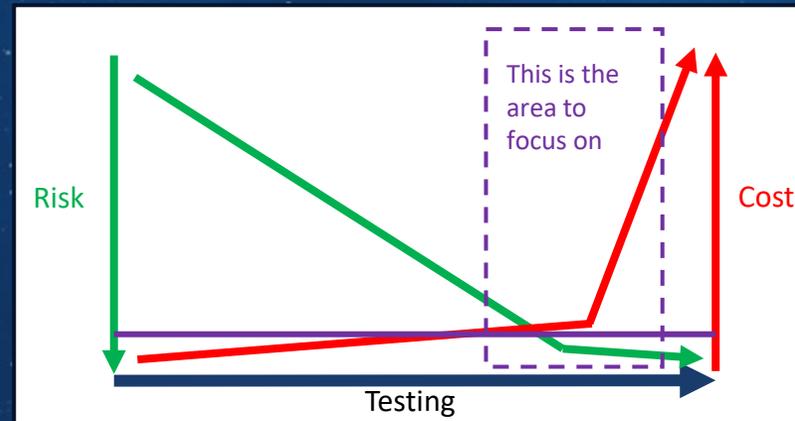
TLYF Operational Scenarios

- The intent of a comprehensive ground test program is to reduce the likelihood of mission impact and crew safety related failures during the mission.
- The term “mission readiness test” has been used in both NASA and non-NASA literature with various specific definitions, but the “gold standard” mission readiness test would be one that occurs just prior to the launch of one or more of the vehicles associated with a mission, using flight hardware and software, exercising all planned (and a select set of off-nominal) ops scenarios, and utilizing flight crew procedures.
- Learning something unexpected about the system, while not a driver for doing these types of tests, is sometimes a valuable unintended consequence, the value of which should be considered in objective trades.



Importance of Proposed TLYF Process

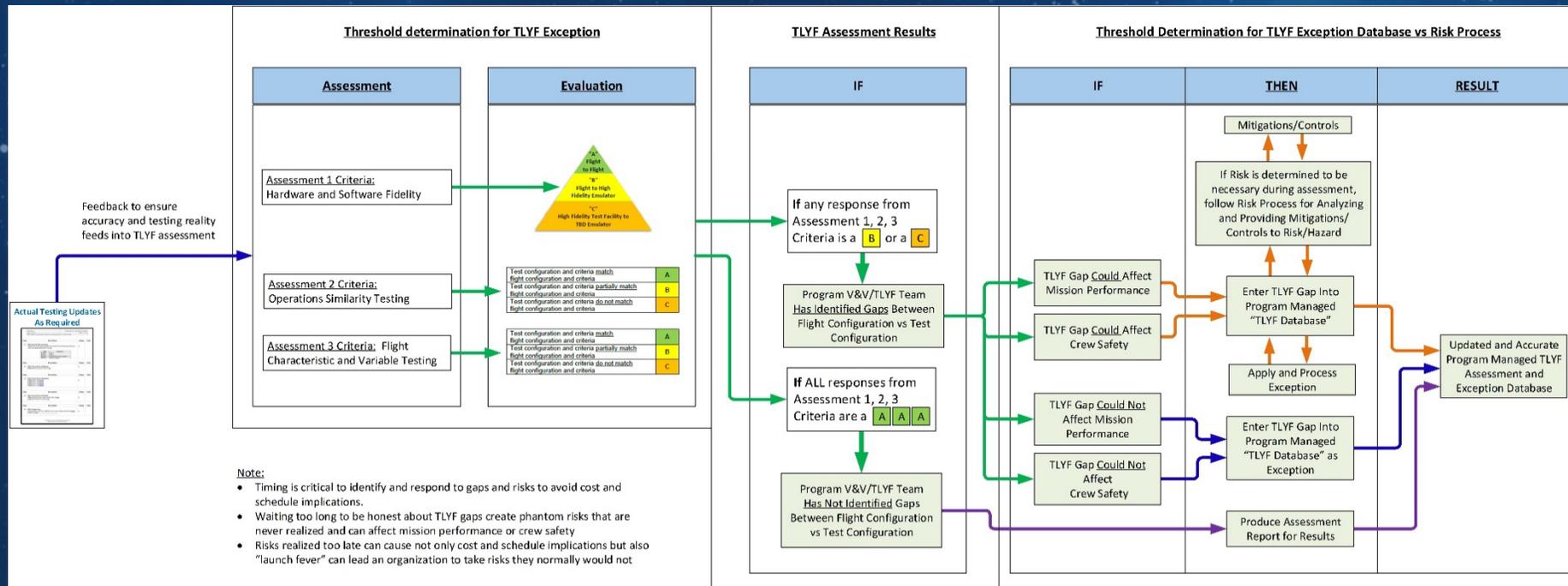
- Risk reduction is central to the TLYF process as it provides a clear goal for why the TLYF process exists and what its products need to provide.
- Processes should always produce useful products that decision makers can use.
- The TLYF process does not replace existing risk processes.
- The TLYF process allows for what is brought into the risk system to be part of an intentional effort to assess gaps between the flight configurations and the test configurations.



Overview of the TLYF Evaluation Process

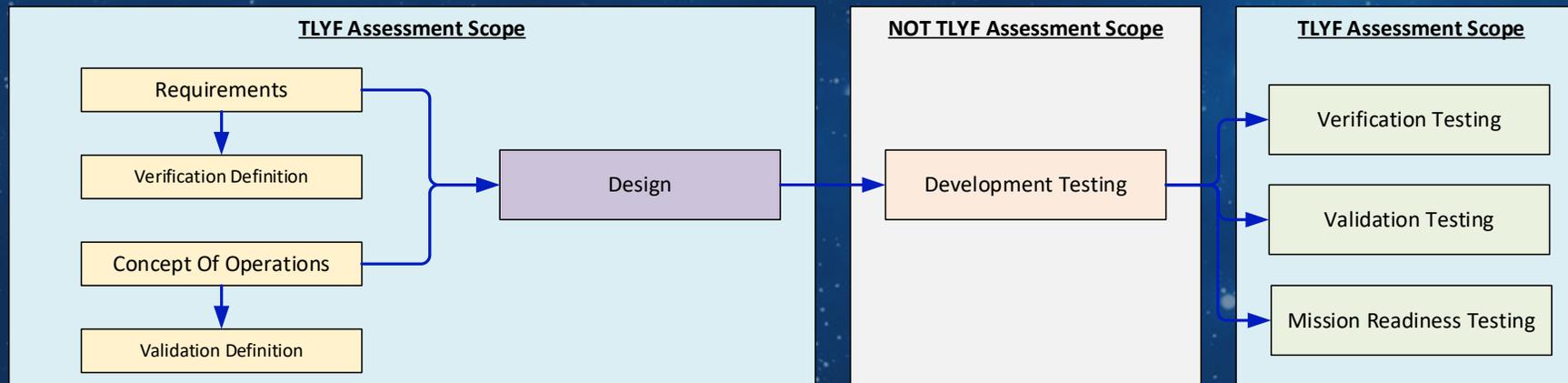
The general steps of the TLYF assessment process flow, illustrated in the figure below are:

- (1) Perform a technical assessment of each planned test, intra-vehicle or vehicle-to-vehicle.
- (2) Create a TLYF exception to document each gap between flight configuration and flight operations verses test configuration and test operations.
- (3) Identify the potential impacts to crew safety and mission success from the TLYF exception.
- (4) Escalate risk to program for evaluation, disposition, and documentation.
- (5) Record assessments and exceptions in a program managed TLYF database.



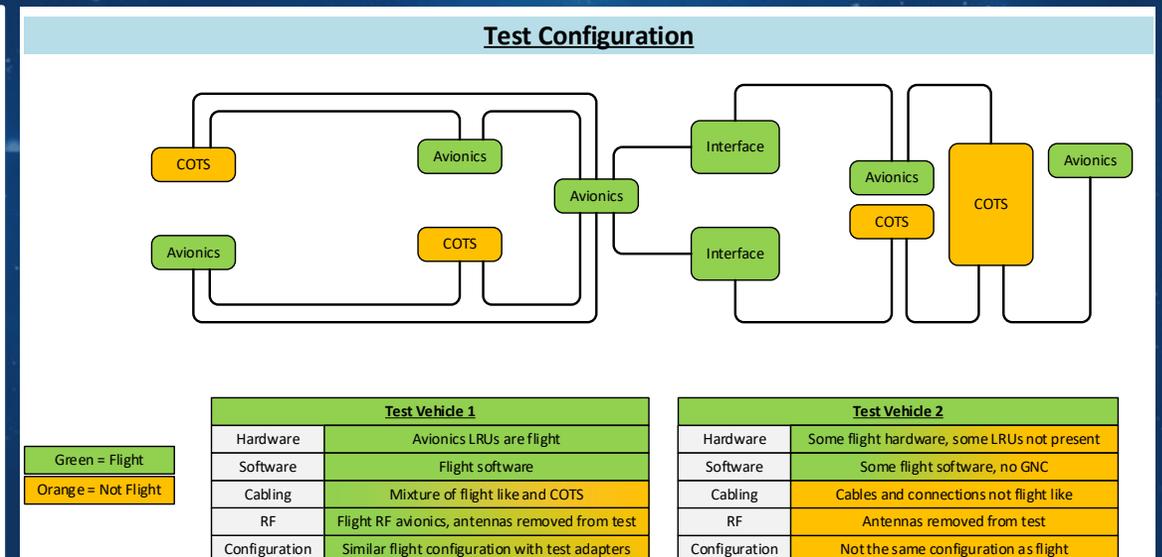
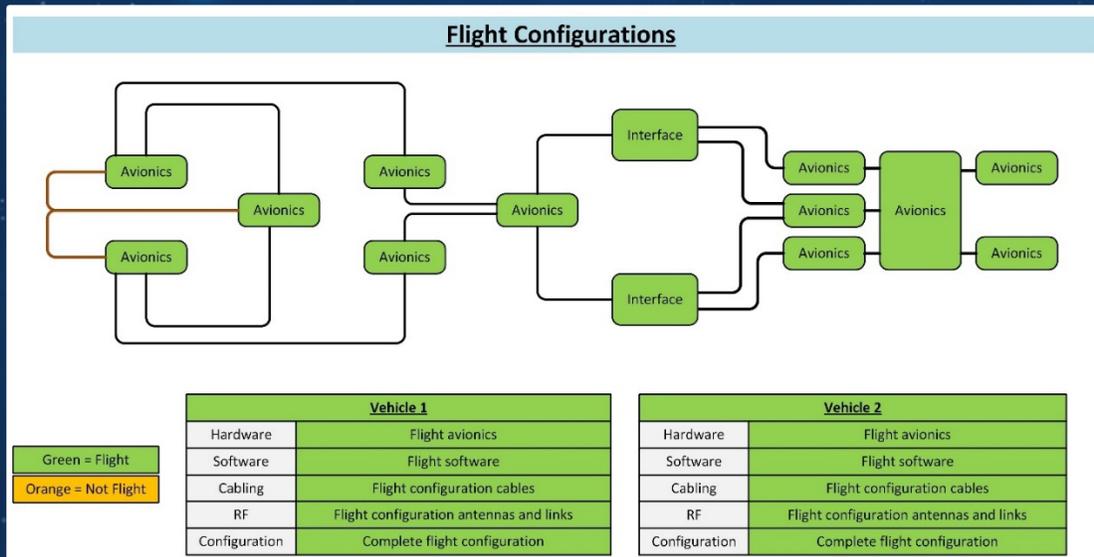
TLYF Process Scope

- What is considered in scope or not in scope for any process is essential to avoid scope creep.
- The TLYF process scope integrates the goals and expectations from the verification and validation plan along with those from any other applicable test plans.
- This HLS TLYF process details the activities that will be performed, the stakeholders, the rubric used for TLYF assessments, and the data used and produced for implementing TLYF products.



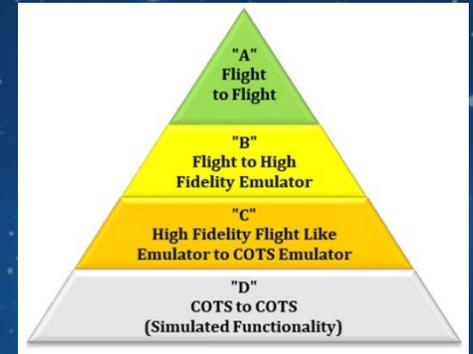
Establishment of Test vs Flight Configuration

- To accurately assess TLYF gaps, it is imperative to clearly understand the flight configuration in all Concept of Operations use cases, configurations and that this is clearly outlined within the TLYF assessment.
- Establishing the source of truth from which everything else is assessed against is critical to success.
- As test objectives are identified and refined and resulting test configurations derived, it is then important to identify not only the hardware to be used for test, but the specific set of flight and ground software products required (and the necessary functions of each) that are needed to accomplish the goals of the test.

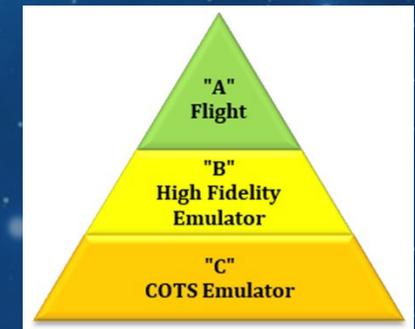


TLYF Criteria Assessment Key

- The flight fidelity of hardware configurations between two elements or within an element under test can be categorized as shown to the right.
- These rubrics allow detailed definitions of the configurations allowing a qualitative assessment approach.
- It is well accepted within the aerospace community that testing against flight hardware allows for the most thorough assessment of the systems performance and mitigation of risks.
- Using hardware/software that is not an accurate representation of the actual flight configuration, or in a test setup that is not reflective of the operational configuration, introduces risk.



TLYF FIDELITY ASSESSMENT ELEMENT TO ELEMENT TESTING



TLYF FIDELITY ASSESSMENT SINGLE ELEMENT TESTING

TLYF Assessment Process

Requirements which are to be verified/validated by testing and demonstration are typical in scope for the TLYF assessment process. As updates occur because of design maturity, controls or mitigations being implemented, the TLYF process will be applied and updated.

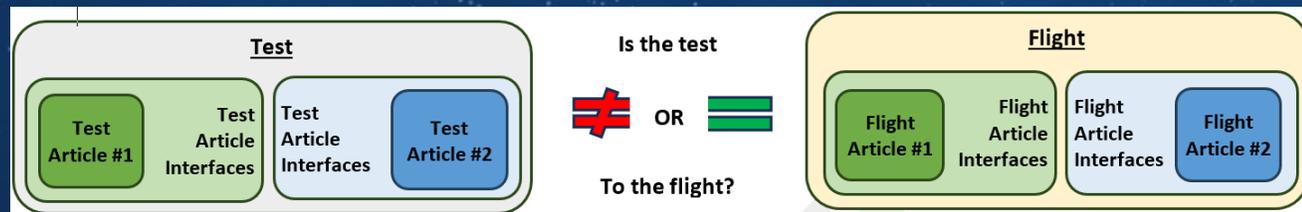
The TLYF Assessment Will:

- 1) Establish the requirement verification evidence and goal of the test
- 2) Establish the flight configuration and root Concept of Operations activity/event
- 1) Establish the test configuration
- 2) Assess hardware similarity
- 3) Assess software similarity
- 4) Assess operational similarity
- 5) Assess characteristics and variables of flight
- 6) Contain a summary of assessment



The TLYF Assessment Then Allows:

- 1) Create a TLYF exception to document each gap between flight configuration and operations and test configuration and operations.
- 2) Identify the potential impacts to crew safety and mission success from the TLYF exception.
- 3) Escalate risk to program for evaluation, disposition, and documentation.
- 4) Record assessments and exceptions in a program managed TLYF database.



TLYF Assessment Criteria #1– Hardware and Software Similarity

- Crucial to a successful TLYF process is to allow the hardware and software lessons learned of the past to be the questions and figures of merit to be used for influencing the future.
- The table below provides for positive recording of specific TLYF hardware and software assessment figures of merit.

TLYF Assessment Criteria #1A– Hardware Similarity

<u>Test Article TLYF Evaluation</u>	<u>Yes/No</u>	<u>Notes/Rationale</u>
1) Is the hardware architecture for Test Article #1 the same as the FLIGHT hardware architecture?		
2) Is the hardware architecture for Test Article #2 the same as the FLIGHT hardware architecture?		
3) Are the interfaces supporting Test Article #1 the same as the FLIGHT hardware interfaces?		
4) Are the interfaces supporting Test Article #2 the same as the FLIGHT hardware interfaces?		
Assessment Criteria 1A Summary: Hardware		
	<u>A/B/C/D</u>	

TLYF Assessment Criteria #1B Software Similarity

<u>Software Category</u>	<u>Software Item</u>	<u>Version</u>	<u>Native/Ported/Development/Model</u>	<u>A/B/C/D</u>	<u>Notes/Rationale</u>
Interfacing Software					
Core Software					
Simulation Software					
Test Scripts					
Software Test Procedures					
Assessment Criteria 1B Summary: Software				<u>A/B/C/D</u>	<u>Notes/Rationale</u>

TLYF Assessment Criteria #2– Operations Similarity

- Crucial to a successful TLYF process is to allow the operational lessons learned of the past to be the questions and figures of merit to be used for influencing the future.
- The table to the right provides for positive recording of specific TLYF operational similarity assessment figures of merit.
- The Evaluation Key provides guidance for evaluation allowing for consistency

Test Capability #1:	A/B/C	Notes/Rationale
Test asset supports all required nominal start to finish mission event and timeline tests.		
Test asset supports all required off-nominal start to finish mission event and timeline tests.		
Test asset supports transitions across operational phases are tested. (e.g. transition from RF to hardline)		
Test Capability #2:		
Test asset will provide fault injection to all required nominal inputs/outputs.		
Test asset will provide fault injection to all required off-nominal inputs/outputs.		
Test assets test each input variable, and the range of each variable is the same as nominal flight conditions.		
Test assets test each input variable, and the range of each variable is the same as off-nominal flight conditions.		
Test Capability #3:		
Test asset uses representative procedure and scripts with TLYF exceptions noted		
Test Capability #4:		
Test specific user input #1		
Test specific user input #2		
Assessment Criteria 2 Summary: Operations	A/B/C	Notes/Rationale

Assessment Evaluation Key:	
Test asset represents and tests <u>all</u> operations, events and procedures	A
Test asset represents and tests <u>some</u> operations, events and procedures	B
Test asset represents and tests <u>no</u> operations, events and procedures	C
This type of variable is not applicable or in scope for this test. This variable should be evaluated by other tests.	N/A

TLYF Assessment Criteria #3– Characteristics and Variables

- Crucial to a successful TLYF process is to allow the detailed technical lessons learned of the past to be the questions and figures of merit to be used for influencing the future.
- The table to the right provides an example of some electrical and RF criteria for positive recording of these specific TLYF technical assessment figures of merit.
- Mechanical, thermal and many other types of technical criteria could be implemented here
- The Evaluation Key provides guidance for evaluation allowing for consistency

<u>Test Assets Accurately Represents These Characteristics & Variables of Flight Configuration</u>	<u>A/B/C</u>	<u>Notes/Rationale</u>
1) Time latencies and constraints are matched with flight conditions		
2) Time responses are matched with flight conditions		
3) Off-nominal transmission characteristic - Power (may have effects on bandwidth)		
4) Off-nominal transmission characteristic - Stochastic Noise (can have effects on bandwidth)		
5) Off-nominal transmission characteristic - EMI/EMC Noise (can have effects on bandwidth)		
6) Off-nominal transmission characteristic - Multipath (can have effects on bandwidth)		
7) RF strings / transmitters represented (including multiple transmitters during joint operations)		
8) Test specific user input #1		
9) Test specific user input #2		

<u>Assessment Evaluation Key:</u>	
Asset represents and tests <u>all</u> characteristics/variables of interface/element	A
Asset represents and tests <u>some</u> characteristics/variables of interface/element	B
Asset represents and tests <u>no</u> characteristics/variables of interface/element	C
This type of variable is not applicable or in scope for this test. This variable should be evaluated by other tests.	N/A

TLYF Exception Application

- A primary goal of the TLYF exception application is to provide clear technical statements that enable stakeholder understanding.
- Equally important is avoiding the normalization of deviance—submission of a TLYF exception application does not guarantee approval and it does not make the risk of the TLYF gap any less. A TLYF exception approval is a program’s acknowledgment of the TLYF gap, understanding of the risk and approval the TLYF gap and exception application is acceptable to the program.
- Clearly communicate what a TLYF exception is and what it means to the mission

~~Writing a lot of words means it looks really great but probably we don't have the best understanding of what it is and all the nuisances clearly but if we keep writing a lot of words over and over hopefully this will make up for not being able to completely understand and communicate risks, actions, solutions, etc but writing a lot of words means it looks really great but probably we don't have the best understanding of what it is and all the nuisances clearly but if we keep writing a lot of words over and over hopefully this will make up for not being able to completely understand and communicate risks, actions, solutions, etc but writing a lot of words means it looks really great but probably we don't have the best understanding of what it is and all the nuisances clearly but if we keep writing a lot of words over and over hopefully this will make up for not being able to completely understand and communicate risks, actions, solutions, etc Writing a lot of words means it looks really great but probably we don't have the best understanding of what it is and all the nuisances clearly but if we keep writing a lot of words over and over hopefully this will make up for not being able to completely understand and communicate risks, actions, solutions, etc but writing a lot of words means it looks really great but probably we don't have the best understanding of what it is and all the nuisances clearly but if we keep writing a lot of words over and over hopefully this will make up for not being able to completely understand and communicate risks, actions, solutions, etc but writing a lot of words means it looks really great but probably we don't have the best understanding of what it is and all the nuisances clearly but if we keep writing a lot of words over and over hopefully this will make up for not being able to completely understand and communicate risks, actions, solutions, etc~~



Clear technical statement
+ Clear risk statement

Clear understanding for stakeholders

TLYF Exception Application Summary Example

- When submitted, the TLYF exception application allows stakeholders to formally determine whether differences between test and flight configurations are acceptable.
- The Test-Like-You-Fly (TLYF) exception application provides a method for stakeholders to acknowledge and approve known differences between test and flight hardware/software/configurations. This application utilizes technical data produced during the TLYF assessment.
- If an application is rejected, this provides necessary traceability and guidance for implementing changes to the test configuration to reduce risks to crew safety and mission success.
- Assessment Evaluation Key:
- The keys below provide insight into the configuration proposed for this test and the evaluated fidelity compared to flight.

Test ID	Title	Assessment Criteria 1: Hardware & Software Fidelity		Assessment Criteria 2: Operations Similarity			Assessment Criteria 3: Characteristics & Parameter Variables	
		Hardware Pedigree (1A)	Software Pedigree (1B)	Test Capability #1	Test Capability #2	Test Capability #3	RF	Hardline
		A/B/C/D	A/B/C/ n/a	A/B/C	A/B/C	A/B/C	A/B/C	A/B/C
TLYF Gap Could/Could Not Affect Crew Safety?		Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No
TLYF Gap Could/Could Not Affect Mission Performance		Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No
Have these components flown previously and successfully in this configuration? (flight heritage)				Yes/No	Rationale			

Assessment Criteria 1A - Hardware:	Grade
Flight Hardware to Flight Hardware	A
Flight Hardware to High Fidelity Emulator	B
High Fidelity Emulator to High Fidelity Emulator	C
COTS Emulator to COTS Emulator – “D”	D

Assessment Criteria 2 - Operations:	Grade
Asset represents and tests <u>all</u> characteristics/procedures/variables of interface	A
Asset represents and tests <u>some</u> characteristics/procedures/variables of interface	B
Asset represents and tests <u>no</u> characteristics/procedures/variables of interface	C

Assessment Criteria 1B - Software:	Grade
Native Software	A
Ported Software	B
Development Software	C
Not Applicable	N/A

Assessment Criteria 3 - Characteristics:	Grade
Asset represents and tests <u>all</u> characteristics/variables of interface	A
Asset represents and tests <u>some</u> characteristics/variables of interface	B
Asset represents and tests <u>no</u> characteristics/variables of interface	C

Final TLYF Recommendations

- The ideal implementation of TLYF principles dictates the use of flight hardware, with flight software and flight operational procedures being tested in the environment the system is intended to operate. To accomplish this best practices effectively, program and project managers must commit, define, approve, and apply this TLYF philosophy as early as possible in the programs or project's life cycle.
- The programs or project's methodical implementation of TLYF principles and approved processes should lead to a test campaign that yields a lower overall residual risk because the highest fidelity hardware, software, and procedures are used and successful completion of the tests indicates that the flight hardware, software, and procedures should operate as expected on-orbit.
- A methodical TLYF process implemented at the beginning of a program/project should yield lower residual risk for flight because it will allow managers to make informed test decisions throughout the life cycle of the program/project when there is time available to make necessary changes. If the TLYF process is only implemented near the end of the program's/project's life cycle, program managers may have to accept a higher level of residual risk because there is no time or capability to correct errors or implement changes.



Thank You