An Integrated Space Test Lexicon

A Taxonomy for the Integrated Test and Evaluation of Space Systems

Stephen K. Tullino
Andrew S. Keys
Robert A. Bettinger
Amy M. Cox
David R. Jacques

The proposed Integrated Space Test Lexicon is intended to amalgamate the numerous definitions of *integrated test* (IT or IT&E), *developmental test* (DT or DT&E), and *operational test* (OT or OT&E) into unified, service-wide definitions, aligned with the *Space Test Enterprise Vision*. Refining such definitions will help distill the core characteristics of these fundamental test types to first identify space system activities composing what is traditionally known as DT and OT, then to provide a means of how these activities fit into the IT paradigm and support space system development. In forging a common understanding of how DT and OT support space systems and capabilities, this lexicon will facilitate the foundation for an IT architecture, specifically the National Space Training and Testing Complex and the larger enterprise-level operational test and training infrastructure.

In March 2022, the US Space Force released its guiding document, *Space Test Enterprise Vision*, where the service laid out its plan of meeting current and future needs. Specifically, the Space Force must integrate operational and developmental space test and evaluation (T&E) activities to meet the challenges posed by the growing threat environment, the rapid emergence of new technologies and capabilities, and the small size of the Space Force. Using the Space Capstone Publication *Spacepower* as guidance, the US Space Force test enterprise strives to address these challenges and "drive data-informed decisions at speed, maximizing the Service's flexibility and efficiency in delivering space-based capabilities for the Joint Force and the nation."[1]

Major Stephen Tullino, USSF, PhD, serves as chief of the Systems Engineering, Integration, and Test Branch for the Operational Test and Training Infrastructure program office, Space Systems Command.

Dr. Andrew Keys is a technology engineer and space area lead for Nou Systems, Inc.

Lieutenant Colonel Robert Bettinger, USAF, PhD, is an assistant professor of astronautical engineering at the Air Force Institute of Technology.

Lieutenant Colonel Amy Cox, USAF, PhD, is the program chair and assistant professor of systems engineering at the Air Force Institute of Technology.

Dr. David Jacques, USAF, Retired, is a professor of systems engineering at the Air Force Institute of Technology.

1. *Space Test Enterprise Vision* (Washington, DC: US Space Force [USSF], 2022), 1, https://www.spaceforce.mil/; and *Spacepower: Doctrine for Space Forces*, Space Capstone Publication (Washington, DC: USSF, 2020).

The Space Force's approach to test and evaluation is centered on the concept of *integrated test* (IT or IT&E)—generally, the consolidation of testing efforts across agencies and the acquisitions cycle. The Space Force will integrate *developmental test* (DT or DT&E) and *operational test* (OT or OT&E) activities—generally, the testing of equipment, munitions, and weapons in the field, and the testing of systems design and performance, respectively—as much as possible "across a capability's life cycle and throughout the test enterprise encompassing organizations, workforce, infrastructure, acquisitions, and operations."[2]

As the number and capabilities of spacefaring countries have increased, the United States must improve its testing to retain its relative advantage across the space domain. Through integrated testing, the Space Force aims to bridge the developmental test-operational test divide by introducing operational perspectives early, simultaneously integrating and enhancing warfighting capabilities.

Considering the lack of consensus on the definitions of these fundamental test types, this article proposes streamlined definitions that align with the *Space Test Enterprise Vision* to identify what are traditionally known as DT and OT and demonstrate how these fit in the IT paradigm and support space system development. Such a lexicon provides a common understanding of how DT and OT support space systems and capabilities within IT. In turn, this understanding will facilitate the foundation for an IT architecture, specifically for the National Space Training and Testing Complex (NSTTC), and the overarching operational test and training infrastructure (OTTI) architecture.

Background

The *Space Test Enterprise Vision* asserts IT "is the collaborative, tailorable, and responsive testing approach to provide shared data for independent evaluation of system performance, effectiveness, suitability, sustainability, and survivability."[3] Testing in the Space Force will be integrated across all levels, both strategic and tactical, from enterprise and system-of-systems level down to a single system and component levels.[4]

Throughout the entire system life cycle—from requirements definition to asset sustainment—testing will involve the individual test professional, who will hold the novel responsibility of providing the resulting IT data to both developmental and operational stakeholders. Testers should be as familiar with programmatic milestones as they are with operational tactics and potential utility. They will be intentionally sourced via workforce crossflow among acquisition, test, and operational professionals, fostering a test culture that promotes Joint warfighter influence upon each system's development and employment.

This crossflow is enabled by the NSTTC, the national network of interconnected, scalable, and distributed range facilities providing realistic threat informed test and

2. *Space Test Enterprise Vision*, 1.
3. *Space Test Enterprise Vision*, 3.
4. *Spacepower*; and Shawn N. Bratton and James P. Seballes, *Vision for: The National Space Test and Training Complex* (Peterson SFB, CO: Space Training and Readiness Command, 2022).

training environments for space warfighters found within the overarching OTTI.[5] The OTTI (fig. 1) is the overarching enterprise-level collection of testing and training assets that includes T&E, operational training, tactics development activities, and the NSTTC. This infrastructure includes blue force devices/trainers/simulators, live/synthetic aggressor capabilities, live and synthetic ranges/environments, and facilities/network that contain and connect OTTI systems.[6]

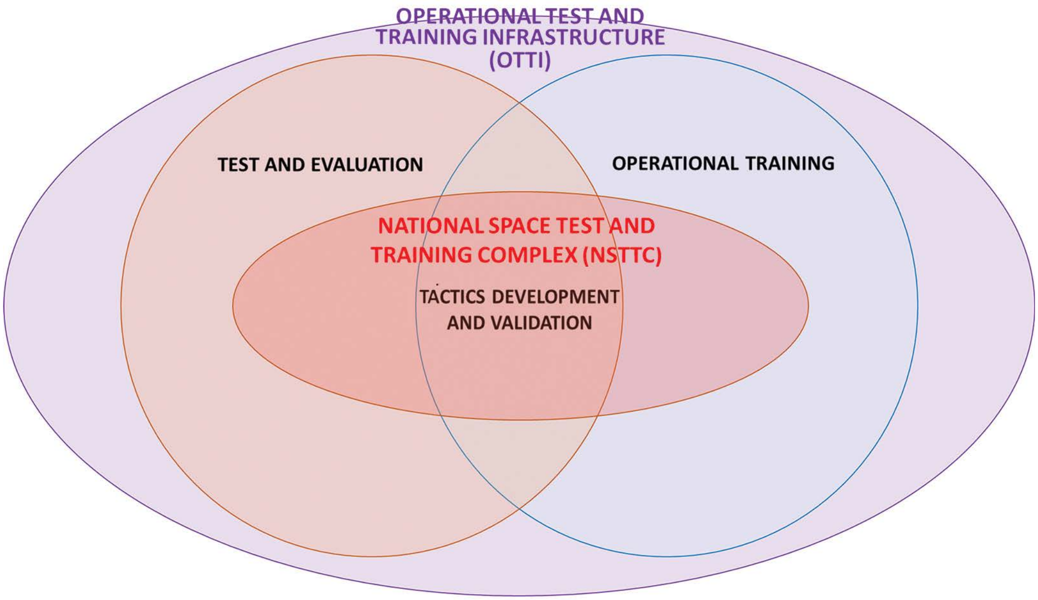

Figure 1. The Operational Test and Training Infrastructure

The terms developmental test and operational test are intentionally used to help bridge the gap between traditional practices and the intent of integrated test per the *Space Test Enterprise Vision*. The proposed lexicon will help develop a common understanding of how DT and OT support space systems and capabilities within IT, and in turn, this understanding will facilitate the foundation for an IT architecture, specifically the NSTTC and OTTI.

Integrated Test Concept

The concept of IT originated via the 1990s acquisition reforms, namely, the 1994 Federal Acquisition Streamlining Act, the 1996 Federal Acquisition Reform Act, and the

5. Bratton and Seballes.
6. USSF/TE, "OTTI vs NSTTC Graphic" (Washington, DC: Department of the Air Force, 2023).

1996 Information Technology Management Reform Act.[7] The goals of the first were threefold: (1) to reduce unique purchasing requirements; (2) to use simplified acquisition procedures for low-income procurement to a greater degree; and (3) to accelerate the acquisition of commercially produced and off-the-shelf services and goods to leverage the latest technologies and reduce the in-house cost of doing business.[8]

The latter two reforms became known as the Clinger-Cohen Act, aiming to implement full and open competition to fulfill government requirements, provide competitive range determinations based on the initial evaluation of proposals versus mandating a certified price for item acquisition, and develop simplified and accelerated procedures for programs under $5 million when expected offers included only commercial items.

According to a 2004 RAND report, all acquisition reform initiatives impacted testing and evaluation specifically by reducing compliance costs and commercial-like practices. These are distilled as requirements reform—operational requirements document flexibility, contractor design flexibility and configuration control via Total System Performance Responsibility, and commercial insertion via commercial-off-the-shelf and nondevelopmental items.[9]

Total System Performance Responsibility intended T&E to emulate the commercial research and development paradigm. Given customer requirements and constraints, contractors had significant control and responsibility over system design, development, and testing. The contractor had primary responsibility for DT design and execution, traditionally seen in spacecraft pre-launch testing, which includes supporting ground systems. A drawback was the unavailability of critical contractor data for government use, which would force the government to recreate and/or purchase contractor data. This responsibility strategy stressed close collaboration with government testers—especially in OT—ensuring user requirements and procedures were fully realized.[10] It led to the combined testing concept: integrating contractor and government DT and OT personnel on a single team, known as the combined test force or integrated test team.[11]

This concept is now known as integrated test. Departing from contractor DT procedurally separated from government OT, IT seeks to combine DT/OT teams to ensure collaboration on ideas and integrate processes from early planning through completion of all major test activities. Benefits include eliminating redundant test activities, early issue resolution, and improved programmatic communication.

Since 2004, several government entities have offered their own IT definitions. For example, the *Integrated T&E Continuum*—as proposed in 2010 by then Director of

7. Michael O'Connell, "Federal Acquisition Streamlining Act (1994)," Federal News Network, 2012, https://federalnewsnetwork.com/; and Defense Acquisition University, Acquipedia, s.v. "Clinger-Cohen Act (CCA)," 2022, https://www.dau.edu/.

8. O'Connell.

9. Bernard Fox et al., *Test and Evaluation Trends and Costs for Aircraft and Guided Weapons* (Santa Monica, CA: RAND Corporation, 2004).

10. Fox et al.

11. Fox et al.

Developmental Test and Evaluation Edward Greer—built upon the Office of the Secretary of Defense's (OSD) IT definition. Under this definition, IT must be an integral part of development and acquisition; efforts or resources between contractor and government DT should not be duplicated—that is, they should be enabled through communication and open sharing of data and knowledge; DT must integrate and flow into OT throughout the acquisitions cycle, as distinct phases progress; and capabilities should be consolidated to ensure sound T&E.[12]

IT facilitates continuous learning and collaboration of knowledge affecting system requirements, development, and performance. Its application across the acquisition life cycle was widely understood and accepted; however, progress was slow due to increased implementation costs; lack of "an interoperable digital engineering environment; interoperability of tooling; and a unified data strategy to support complete end-to-end knowledge and data sharing throughout the T&E phases and acquisition lifecycle."[13]

The current IT definition has been criticized because it omits an emphasis on the criticality of information and data throughout the life cycle.[14] This definition of IT was to respect independence, but it has become an obstacle, especially in space with respect to duplication of effort and/or paying for data. Authoritative sources show inconsistent definitions, challenging effective IT practice and implementation.

Following similar efforts, this article uses a democratic approach—where different sources are analyzed to identify and extract common themes and concepts—to consolidate a definition.[15] One scholar argues language is the "accepted method of human communication" to foster understanding. A precise use of language is preferred because it stresses accurately using words to achieve consensus.[16] Table 1 analyzes authoritative definitions of IT and demonstrates how the democratic approach works. Column headers show the concepts most identified in analyses. These concepts include unified team, involved team—an entire team involved at all acquisition stages—and data-sharing.

---

12. Edward R. Greer, "Developmental Test and Evaluation Is Back," *International Test and Evaluation Association* [ITEA] *Journal* 31 (2010), https://apps.dtic.mil/; and Charles E. McQueary and James I. Finley, Office of the Secretary of Defense, Memorandum for Component Acquisition Executives, Subject: Definition of Integrated Testing, April 25, 2008, https://www.dote.osd.mil/.

13. Laura Freeman, Geoffrey Kerr, and Jeremy Werner, "Positioning Test and Evaluation for the Digital Paradigm Building Blocks for the DE Transformation," *Journal of Test & Evaluation* 44, no. 2 (2023).

14. Stephen Tullino, interview with Dr. Andrew Freeborn, United States Air Force Test Pilot School Space Test Fundamentals Course Director, Edwards AFB, CA, May 4, 2023; and Freeman, Kerr, and Werner.

15. Jeremy R. Geiger, "Agility Measurement for Large Organizations" (PhD dissertation, Air Force Institute of Technology, September 2020), 7, https://scholar.afit.edu/; and Erin T. Ryan, David R. Jacques, and John M. Colombi, "An Ontological Framework for Clarifying Flexibility-Related Terminology via Literature Survey," *Systems Engineering* 16, no. 1 (2012), https://doi.org/.

16. Ryan, Jacques, and Colombi.

Table 1. Summary of integrated test definitions

Source	Definition	Unified	Involved	Data
OSD Guidance Memorandum (2008)	"collaborative planning and collaborative execution . . . provide shared data . . . by all stakeholders"[17]	–	X	X
DoDI 5000.89 *Test and Evaluation* (2020)	"capitalizes on the idea that test events can . . . provide data for . . . evaluations."[18]	–	–	X
Director, Operational Test and Evaluation (DOT&E) Test and Evaluation Master Plan (TEMP) Guidebook vers. 3.1 (2017)	"Developmental Test incorporates characteristics of Operational Testing . . . or the data from Developmental Testing is accepted as adequate for the operational evaluation"[19]	–	–	X
Test and Evaluation Enterprise Guidebook (2022)	"merge the primary test stakeholders . . . into one unified test team . . . data sharing among all . . . utilizing all test events . . . in the program to achieve CT, DT, and OT objectives in a collaborative fashion to the maximum extent possible"[20]	X	X	X
DAFI 99-103 *Capabilities-Based Test and Evaluation* (2022)	"collaborative planning and execution of test phases and events to provide shared data . . . by all [DT & OT] stakeholders"[21]	–	X	X
DAFMAN 63-119 *Mission-Oriented Test Readiness Certification* (2021)	"collaborative planning and execution of test phases . . . to provide shared data . . . by all [DT & OT] stakeholders"[22]	–	X	X
SECNAV Instruction 5000.2G *Department of the Navy Implementation of the Defense Acquisition System and the Adaptive Acquisition Framework* (2022)	"Programs should . . . allow the DT and OT communities to gather needed information in the proper environment/conditions as early as possible . . . to inform programs . . . potentially reducing the scope of dedicated OT events. . . . The goal of an integrated testing event is to ensure that the collected data will be usable for DT and OT."[23]	–	X	X

17. McQueary and Finley, Memorandum.

18. Office of the Under Secretary of Defense for Research and Engineering, and Office of the Director, Operational Test and Evaluation (DOT&E), *Test and Evaluation*, Department of Defense (DoD) Instruction 5000.89 (Washington DC: US DoD, 2020), 38, https://www.dote.osd.mil/.

19. J. Michael Gilmore, *DOT&E Test and Evaluation Master Plan (TEMP) Guidebook*, vers. 3.1 (Washington, DC: Office of the Secretary of Defense, January 9, 2017), 101, https://www.dote.osd.mil/.

20. Heidi Shyu and Nickolas H. Guertin, *Test and Evaluation Enterprise Guidebook* (Washington, DC: Under Secretary of Defense for Research and Engineering, and DOT&E, 2022), 6–7, https://www.test-evaluation.osd.mil/.

21. *Capabilities-Based Test and Evaluation*, Department of the Air Force (DAF) Instruction 99-103, DoDI 5000.89 (Washington, DC: DAF, November 9, 2021, corrective actions applied on March 15, 2022), 37, https://static.e-publishing.af.mil/production/.

22. *Mission-Oriented Test Readiness Certification*, DAF Manual 63-119 (Washington, DC: DAF, April 15, 2021), 35, https://static.e-publishing.af.mil/.

23. Carlos del Toro, *Department of the Navy Implementation of the Defense Acquisition System and the Adaptive Acquisition Framework*, Secretary of the Navy (SECNAV) Instruction 5000.2G (Washington, DC: Department of the Navy, 2022), 23, https://www.secnav.navy.mil/.

Source	Definition	Unified	Involved	Data
Space Test Enterprise Vision (2022)	"Integration of effort across the spectrum of test activities . . . fosters early operational input into system development, while providing the developer with greater insight into the intended employment of the system."[24]	–	X	X
USSF/TE Space Test Enterprise Brief (2022)	"A collaborative, tailorable, and responsive testing approach to provide shared data for independent evaluation of system performance, effectiveness, suitability, sustainability, and survivability."[25]	X	–	X
USAF Arnold Engineering Development Center (2004)	"integration of modeling tools, including computations and engineering methods, in direct support of ground and flight tests"[26]	–	–	X
Delta 12 Test Guidebook	"IT combines DT and OT events to achieve greater efficiency, reduce cost, and/or accelerate . . . without compromising . . . objectives . . . Integrated Testing requires the collaborative planning and execution . . . to provide shared data in support of . . . all stakeholders. Whenever feasible, T&E campaigns will be conducted in an integrated fashion to permit all stakeholders to use data in support of their respective functions . . . early engagement with program offices and staff, test teams can ensure that system requirements are testable and T&E requirements are meaningful. IT enables early identification of system design issues and guides the system development"[27]	X	X	X
Air Force Test and Evaluation Guide (2019)	"Integrated Testing in operationally representative environments is the best method to understand performance of complex systems. Programs can accelerate learning . . . by conducting early mission-focused testing in relevant environments utilizing . . . the most stressful combinations or most likely use cases. This strategy can also expose potential operational issues early . . . and [reduce] time-consuming delay[s] towards the end of a program's development."[28]	–	X	X
RAND Corporation's *Test and Evaluation Trends and Costs for Aircraft and Guided Weapons* (2004)	"integration of . . . personnel on a single test team . . . [who] are involved from the early planning stages through . . . completion of all major test activities . . . early involvement of OT personnel in DT saved both costs and schedule"[29]	X	X	X
TOTAL (Out of 13)		4	9	13

Table 1 reveals the different definitions' intended contexts and the extent to which they represent a concept of an operationally unified team, an involved team, and data

24. *Space Test Enterprise Vision*, 7.
25. *USSF Space Test Enterprise Vision Brief* (Washington, DC: Headquarters, USSF, September 2022), 9.
26. Marcus L. Skelley, Tommie F. Langham, and William L. Peters, "Integrated Test and Evaluation for the 21st Century," 3, (paper, USAF Developmental Test and Evaluation Summit, Woodland Hills, CA, November 16–18, 2004), https://doi.org/.
27. *Delta 12 Test Guidebook (Working Copy)* (Peterson SFB, CO: Space Delta 12, 2021), appendix A.
28. *Air Force Test and Evaluation Guide Combined v. 2* (Washington, DC: USAF Test & Evaluation, September 24, 2020), 52, https://www.dau.edu/.
29. Fox et al., *Test and Evaluation Trends*, 46.

sharing throughout the test campaign. Four definitions included a unified team, nine included an involved team, and all 13 included data sharing. The terms *unified* and *involved* could be consolidated into the term *collaborative*, more accurately reflecting OSD verbiage. Using this verbiage, 10 definitions would include collaborative. These yield the following proposed definition:

> Integrated test is the streamlining and consolidation of system test activities and datapoints via collaborative planning, collaboration, and support among government and contractor agencies throughout the entire acquisitions cycle. Its objective is to verify that the system meets specifications and is validated to use cases to meet mission needs. This includes assessing combat capabilities and integration within the Joint warfighting construct. Shared data from iterative test activities are leveraged to gain the following advantages: reduce waste and risk, improve system design and performance, and increase communication among program and test management teams.

Interpretation error risks are reduced by evaluating each definition within its original context. For example, some sources define IT through lessons learned and best practices.[30] With this synthesized IT definition, this method is repeated for DT and OT.

Developmental Test for Space Systems

Per DoD Instruction (DoDI) 5000.89, *Test and Evaluation*, developmental test informs decisionmakers, characterizes and troubleshoots system designs, matures technology via risk reduction, and prepares for OT.[31] Experts from the US Air Force Arnold Engineering Development Center add, "Along with flight testing, the ultimate goal of [DT] is to produce a complete picture of an optimized flight vehicle."[32] A RAND report further scopes this where "DT is performed at the part, subsystem, or full system level to prove design validity or reliability, materials used, etc. . . . [and] results are used to modify the system design to ensure that it meets the design parameters and system specifications."[33]

DT ensures developing technology and systems meet key performance parameters; the program office ensures this via increasing technology readiness levels (TRLs) from TRL 4 to TRL 7, known as the "Valley of Death."[34] Space Systems Command is the Space Force program office and organization tasked with DT by the *Space Test Enterprise Vision*. In 2017, Space Systems Command's chief scientist noted DT was "focused on meeting

30. Geiger, "Agility Measurement."
31. DODI 5000.89; and DAFI 99-103.
32. Skelley, Langham, and Peters, "Integrated Test and Evaluation," 11.
33. Fox et al., *Test and Evaluation Trends*, 13.
34. Elozor Plotke, Peter C. Lai, and Roberta M. Ewart, "Using Small Satellites to Construct an In-Space Test Platform for Risk Reduction," AIAA SciTech 2023 Forum, National Harbor, MD, January 23–27, 2023, https://doi.org/; and Marshall Smith et al., "Free-Flying StarLabs as Platforms for InSpace Developmental Test," AIAA SciTech Forum 2022, San Diego, CA, January 3–7, 2022, https://doi.org/.

detailed technical specifications . . . normally conducted with the contractor" and emphasized that such testing was traditionally designed to evaluate a system prior to a threat's presence via reliability, availability, and maintainability.[35]

Though well defined, there is much debate on what developmental test means for assets deployed in space versus air domains. As done with IT, the democratic approach enables the identification of what DT means for space systems by extracting key aspects via nine distinct definitions stipulated by 11 authoritative sources, with some duplicating others:

- DoD Instruction 5000.89;
- *Capabilities-Based Test and Evaluation*, Department of the Air Force (DAF) Instruction 99-103;
- *Test and Evaluation Policy*, Army Regulation 73-1;[36]
- *Implementation of the Defense Acquisition System and the Adaptive Acquisition Framework,* Secretary of the Navy Instruction 5000.2G;
- *Test and Evaluation Enterprise Guidebook*;
- *Mission-Oriented Test Readiness Certification,* DAF Manual 63-119*;*
- *Air Force Test and Evaluation Guide;*
- *Delta 12 Test Guidebook;*
- *Combined SMC T&E Guidebook* (2019);[37]
- Ewart's "SPACE Cyber Test and Evaluation Strategies for Space Enterprise Vision";
- RAND Corporation's *Test and Evaluation Trends and Costs for Aircraft and Guided Weapons* (2004).

From these, common attributes consistently emerge, listed in order of frequency: evaluate design, technical readiness, programmatic readiness, characterize systems, OT readiness, relevant environment capability, and finally, contractor involvement. Further, half of the documents do not mention "relevant environment capability" and "contractor involvement." Contractors perform most DT for space assets; there are some DAF units that conduct DT.[38]

35. Roberta M. Ewart, "SPACE Cyber Test and Evaluation Strategies for Space Enterprise Vision," AIAA SPACE and Astronautics Forum and Exposition, September 12–14, 2017, https://doi.org/.

36. *Test and Evaluation Policy*, Army Regulation 73-1 (Washington, DC: Headquarters, Department of the Army, June 8, 2018), https://armypubs.army.mil/.

37. Engineering Directorate – Test & Evaluation Branch, *Test & Evaluation Guidebook – Combined Guidebook* (Los Angeles AFB, CA: Space and Missile Systems Center, 2019), 33.

38. Committee on Armed Services, National Defense Authorization Act for Fiscal Year 2013, "Air Force Space Developmental Test and Evaluation," S. Rep. 112-173, 59, https://www.congress.gov/.

A 2004 RAND report on T&E trends for aircraft and weapons noted the contractor conducted representative environmental/design tests as part of DT—including, but not limited to, modelling and simulation, wind tunnel tests, static article tests, avionics integration tests, special test articles, ground tests, armament/weapon delivery integration tests, and system test requirements and planning.[39]

These activities noted by RAND provide a comparable baseline to identify the essence of spacecraft DT, specifically contractor-run environmental and functional tests at unit, subsystem, and integrated levels. The Aerospace Corporation's *Space Vehicle T&E Handbook* notes tests accomplished by the contractor for a mission at the unit, subsystem, and integrated system levels typically include acoustic, vibration, shock, thermal vacuum, thermal cycling, and electromagnetic interference/charge—which evaluate the system's capability of surviving the harsh launch and space weather environments.[40]

A common methodology is iterative verification and validation (V&V) of units/subsystems/systems via "in-the-loop" means where a system is initially modelled algorithmically (algorithm-in-the loop), digitally emulated using its intended software (software-in-the-loop, SIL), and using hardware in a test stand to verify interfaces and controllability (hardware-in-the-loop, HIL).[41] Hardware-in-the-loop testing is used in the V&V of unit, subsystem, and preliminary integrated system testing. Traditionally, in spacecraft testing, the final iteration of tests can be and often are considered OT, as they are the final gamut of ground tests conducted before the spacecraft is declared cleared for launch and operations.

Given these analyses, the proposed distilled definition for space DT is:

> Developmental test and evaluation verifies that the system meets specified programmatic and technical requirements and specifications via the assessment and characterization of a system's technical performance, reliability, and maintainability. Contractors and government personnel iteratively conduct DT tests to ensure the system is technologically mature, with the technical readiness for operational testing—whether on the ground or in space. DT data feedback is shared to improve system design and performance, and to inform all involved program and test management teams.
>
> In space systems, DT is achieved through modelling and simulation activities and environmental, safety, and functional—hardware and software in-the-loop—testing at component, subsystem, and when necessary, system levels.

39. Fox et al., *Test and Evaluation Trends*.

40. National Systems Group, *Space Vehicle Test and Evaluation Handbook*, ed. J. D. White, G. A. Larsen, and D. W. Hanifen, 2nd ed. (El Segundo, CA: Aerospace Corporation, 2012).

41. Jens Eickhoff, *Simulating Spacecraft Systems* (Heidelberg, Germany: Springer, 2009), https://doi.org/.

Operational Test for Space Systems

Unlike DT, operational test is explicitly defined by law:

> The field test, under realistic combat conditions, of any item of (or key component of) weapons, equipment, or munitions for the purpose of determining the effectiveness and suitability of the weapons, equipment, or munitions for use in combat by typical military users; and the evaluation of the results of such test.[42]

Operational test is not a series of mandatory, independent operator tests, or an inspection or duplication of DT. Developmental test stresses specifications to identify engineering and design deficiencies where designers, contractors, and scientists evaluate the system under test in controlled, often ideal, conditions. Operational test is more focused on user/operator needs where the system under test is evaluated in stressed and operationally realistic scenarios, typically by the operating command's personnel with normal operations and maintenance skills.[43] This type of test is legally straightforward, but its nuances still must be determined to better comprehend what OT is meant for spacecraft.

Again, elements of OT from the same authoritative sources as DT are extracted via the democratic approach, and the following common themes emerge repeatedly, in order of frequency: realistic operations conditions, effectiveness, suitability, tactics, techniques, and procedures development, threat survivability/resiliency, and programmatic readiness.

Testing and evaluation are crucial to ensure a system is programmatically ready to be operationally fielded. For space systems, OT—especially for threat survivability/resiliency—has been challenging. In 1981, the since-renamed Air Force Operational Test and Evaluation Center noted major space systems need to undergo "purposeful and significant" OT.[44] The long space system design and development timelines require OT's impact be felt early in the acquisitions cycle to influence final design and production. In traditional acquisitions, OT occurs before the production or major investment decisions, while most investments for space systems occur early, often without a major production decision.

One study noted this was still problematic in 2008, where space OT agencies tested via a model designed for large-scale production systems not appropriate for space testing.[45] Traditional programs have low expenditures in research and development compared to production and system operations, while space systems experience the opposite. Operational tests still must represent an assessment of the actual operational space system since such tests require an environment as realistic as possible, typically done via (1) data analyses of DT events, (2) a test in representative environments, or (3) a test in space.

42. 10 U.S.C. § 139 (2011).

43. Patricia Sanders, "Challenge for Today: Operational Test and Evaluation of Space Systems," 1st Flight Test Conference, Las Vegas, NV, November 11–13, 1981, https://doi.org/.

44. Sanders.

45. Stephen T. Sargeant and Suzanne M. Beers, "AFOTEC's Space Test Initiative: Transforming Operational Testing and Evaluation of Space System Capabilities," *ITEA Journal* 1, no. 29 (2008), https://apps.dtic.mil/.

Operational test objectives using DT data are met through analyzing procedural data, factory/lab tests, ground processing, and flight operations/data. These test measures are often injected into DT events to fulfill OT objectives. Objectives trace mission needs, operational requirements, system specifications, previous test results and experiences, and the Test and Evaluation Master Plan.[46]

The Global Positioning System and Defense Support Program missions conduct this process via independent evaluation teams, applying OT objectives to contractor testing and DT data. Launch range compatibility testing does this during satellite-booster mating and OT test measures, and scenarios are inserted into DT checkout/compatibility tests, yielding operational impact assessments.[47] These are specific examples of spacecraft IT because the spacecraft's unique exposure to risks—due to the harsh environments of launch, zero gravity, extreme temperatures, and orbital maneuvering—dictate that testing activities and data points be integrated as much as possible. Data can also be collected from similar space vehicle subsystems' reliability statistics.

Satellites are normally declared operational after successful launch and early operations (LEOPS) checkouts and after stably operating for a determined time. Such operations are the most critical due to the high-risk launch environment and the full commissioning of the spacecraft.[48] In 2008, the Air Force Operational Test and Evaluation Center noted OT of space systems still had occurred after launch and ground stations were fielded, resulting in the inability to provide timely and independent OT data to decisionmakers. This compounded the problem posed by the fact that, as noted above, space system investments and decisions occurred early in the program, often without OT data.[49]

Flight environment OT defining the "flight/performance envelope" is difficult because it requires excessive commands and mandates straining a system's redundant capabilities, which are programmatically considered "high risk endeavors."[50] On-orbit OT is difficult to discreetly conduct out of the view of adversaries and poses risks to military missions. This necessitates modeling and simulation (M&S), as well as leveraging OT in other ways during DT or additional on-ground operational testing. Because of this, LEOPS deployments, telemetry monitoring, and checkouts are currently the pinnacle events of space OT.

Space Delta 12 is the Space Training and Readiness Command (STARCOM) organization responsible for US Space Force OT. The space delta executes space system OT in operational scenarios, which help define a spacecraft's performance envelope—akin to aircraft methods—when conducting operations in nominal, natural and hostile, and intentional

---

46. Sargeant and Beers; and Sanders, "Challenge."
47. Sargeant and Beers.
48. James R. Wertz, David F. Everett, and Jeffery J. Puschell, eds., *Space Mission Engineering - The New SMAD* (Hawthorne, CA: Microcosm Press, 2015).
49. Sargeant and Beers, "AFOTEC's Space Test Initiative."
50. Sanders, "Challenge."

threat conditions.[51] Traditionally, testing mostly evaluates the spacecraft's hardiness against environmental and launch stresses. Evaluating a spacecraft in hostile environments requires further tests against various threats. In 1981, the Air Force Test and Evaluation Center noted that a realistic testing environment can only be partially attained through simulating conditions to test subsystems, providing limited scope in establishing confidence.[52]

Given this, spacecraft OT can be distilled into the following definition:

> Operational test and evaluation focuses on validating whether the system meets operational user and mission needs, and whether it can be employed in operational use cases—both nominal and contested scenarios—in the intended way. It does this by evaluating a system's effectiveness, suitability, and survivability in realistic operational conditions.
>
> In spacecraft, OT is typically conducted through (sub)system verification on non-developmental test environmental tests—thermal, radioactive, etc.—verification of DT data, concurrent execution of OT data points during contractor DT events, and on-orbit tests during early and full operations.

Integrated Test for Space Systems

Space system T&E space environment difficulties forced the space community to conduct rudimentary IT via concurrently executing OT datapoints during contractor DT events—that is, incorporating OT characteristics into DT events—and verifying and accepting DT data for operational evaluation. This correlates with the *Director, Operational Test and Evaluation (DOT&E) Test and Evaluation Master Plan (TEMP) Guidebook* as follows.[53]

Characterizing and measuring capacities not reliant on test conditions can be satisfied via DT and can be included in OT evaluations. Testing must use engineering development units and/or models and testing production representatives to measure performance and characterize (sub)system performance as well as hardware-in-the-loop tests. Testing must also use environmental test data to collect OT datapoints as well as historical data and testing of subsystems in thermal, electromagnetic, and radiation environmental tests, as mentioned before. Development test events should be conducted under sufficiently operationally realistic conditions.

In LEOPS and normal operations all space-based tests can be observed, making testing difficult. Modeling and simulation via SIL/HIL and other computer models can bridge this gap and reduce risk in small spacecraft (SmallSat) missions.

---

51. *Space Test Enterprise Vision*; and Stephen Tullino, interview with Colonel E. Lincoln Bonner, commander, Space Delta 12, Schriever AFB, CO, March 24, 2023.

52. Sanders, "Challenge."

53. Gilmore, *DOT&E TEMP*.

Schedule and resource limitations traditionally forced SmallSats and CubeSats—a class of nanosatellites—to be solely functionally verified through simulations.[54] These missions had low success rates, approximately 48 percent in 2010. Until 2015, 20 percent of CubeSats failed post-launch.[55] Recent missions began to use HIL techniques used in traditional satellite testing via flat satellites (FlatSats) for (sub)system V&V.

Using FlatSats for CubeSats can be considered an example of IT because many of these programs iteratively test their satellites as designs mature. This lean approach is necessary due to smaller teams, less funding, and shorter timelines.[56] FlatSats are powerful because they enable rapid flaw identification and correction in design, interfacing, and most nonmechanical issues. FlatSats allow early functional test development, enabling early identification of software and hardware design flaws, so that flight hardware testing focuses only on workmanship.

The feedback-intensive nature of FlatSats ensures the robustness of the system design and software. This is used at the Air Force Research Laboratory SmallSat Portfolio, where its satellites undergo full system V&V through four critical tests: (1) a command and execution test in which full software functionality executes each command; (2) a power characterization test which tests full power subsystem functionality, characterization, and safety limits; (3) a long-range communications verification test ensuring communication links close; and (4) a day-in-the-life test that demonstrates spacecraft critical functionality and all operational modes. This final test ensures the system performs as intended and executes the LEOPS sequence. [57]

These tests are iteratively conducted via emulator (SIL), HIL FlatSat with engineering units, FlatSat with flight units, and fully integrated spacecraft. Each instance of these tests identifies problems and informs the subsequent iteration up to the point where the LEOPS sequence during spacecraft commissioning is executed almost out of routine, in line with Test Like You Fly. Iterative in-the-loop methods qualify and characterize the system in a timely manner, while increasing design robustness and capability confidence—a powerful tool capable of crossing DT, OT, and M&S. Regarding Test Like You Fly, STARCOM is tasked with establishing a network of ranges focused on providing realistic threat-informed test and training environments known as the NSTTC.[58]

This complex focuses on four areas: service capability, Joint applicability, IT, and threat replication. These tests are supported via electromagnetic, on-orbit, cyber, digital, and multidomain command and control. The NSTTC integrates multiple venues, leveraging

54. Sabrina Corpino and Fabrizio Stesina, "Verification of a CubeSat via Hardware-in-the-Loop Simulation," *IEEE Transactions on Aerospace and Electronic Systems* 50, no. 4 (2014), https://doi.org/.

55. João Cláudio Elsen Barcellos et al., "FlatSat Platforms for Small Satellites: A Systematic Mapping and Classification," *IEEE Journal on Miniaturization for Air and Space Systems* 4, no. 2 (2023), https://doi.org/.

56. Jared Clements et al., "Tailored Systems Engineering Processes for Low-Cost High-Risk Missions," in *Space Education and Strategic Applications* 1, no. 1 (2019), https://doi.org/.

57. Clements et al.

58. Bratton and Seballes, "Vision."

on-orbit, digital, HIL, lab, and chamber testing—supporting one another in achieving test objectives. Testing and training focus on multilevel blue-force M&S, from digital twins to exquisite capabilities; program validation via integrating DT, OT, and training; and tactics, techniques, and procedures validation. These align to the in-the-loop test approach as an effective means to marry developmental test, operational test, and modeling and simulation.

Incorporating these, the guiding IT philosophy for combat space systems has been defined as "the use of test to learn (i.e., characterize) as much about the combat capabilities of space systems as practical at all times, regardless of system maturity."[59] The integrated test framework (fig. 2) elaborates on this.

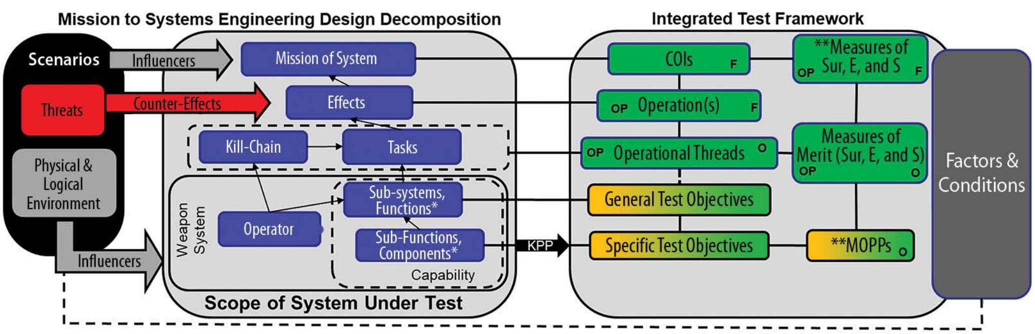

Figure 2. STARCOM Delta 12/4th Test and Evaluation Squadron mission-to-systems engineering and IT framework[60]

Guided by survivability, susceptibility, vulnerability, and recoverability, the IT framework decomposes mission needs to flexibly meet objectives with respect to factors and conditions. The IT framework introduces one new concept, *measures of merit*. Delta 12 posits these to bridge the gap between mission and functionality—in between measures of effectiveness and performance. These are focal points of analysis to describe system performance of a task, or an aspect related to how the mission was executed. This is an additional tier proposed by Delta 12 and permitted under STARCOM guidance.

The IT framework ties to systems engineering design (fig. 2), as the framework translates the engineering design into characterizable (nonrequirements) and evaluable (requirements) aspects against operational scenarios determining how well the system under test performs its mission in its intended way. This includes any new systems or aspects, operational tactics, tasks or kill chains for evaluation, and enhancements to the test framework and systems engineering designs.

59. E. Lincoln Bonner, "Space Delta 12 Update and Integrated Test Force (ITF) Construct" (presentation, 2023 Space Test and Evaluation Summit, Colorado Springs, CO, 2023).

60. Kenneth H. Carpenter III, *Test Combat Framework Guide* (Schriever SFB, CO: STARCOM Delta 12, 4th Test and Evaluation Squadron, 2023).

This framework manifests in the Integrated Survivability Assessment as inspired by the *DOT&E TEMP Guidebook*, where operational scenarios guide the design systems' DT so that test data can naturally inform OT parameters.[61] This can be adapted to show how IT could be carried out on a system (table 2).

Table 2. Adaptation of integrated assessment paradigm to space system T&E

	DT	M&S	OT / Live Fire
Mission Planning System	X (hardware-in-the-loop [HIL])	X (algorithm-in-the-loop [AIL], software-in-the-loop [SIL], HIL)	X (HIL, Spacecraft)
TTPs	X (HIL)	X (AIL, SIL, HIL)	X (HIL, Spacecraft)
Spacecraft Signatures (RF, IR, Visual)	X (Unit Testing, Environmental)	X (Representative Model)	–
Spacecraft Performance Envelope	X (HIL, Environmental)	–	–
Software	X (AIL, SIL, HIL, Spacecraft)	X (AIL, SIL, HIL)	–
Sensors Envelopes	X (Unit Testing w/ HIL, Environmental)	–	–
Subsystems Envelopes	X (Unit Testing w/ HIL, Environmental)	–	–
Threat Tolerance (Vulnerability)	–	X (Representative Model)	X (Environmental & On-Orbit)

Test objectives can be arranged via what is traditionally DT, OT, and M&S, complementing and improving upon each other and/or enabling a more holistic yet efficient approach, especially through in-the-loop means. Model V&V is iteratively achieved—data being checked against itself and against specifications—and to intended use cases, using realistic operational data where possible.

Conclusion

Existing developmental test, operational test, and integrated test terminology suffer from ambiguities and inconsistencies, leaving testers to guess the scope of these test activities and thus creating challenges in meeting the spirit and intent of the *Space Test Enterprise Vision*.[62] To ensure US Space Force field commands consistently adhere to this

61. Tullino, interview with Bonner.
62. *Space Test Enterprise Vision*.

vision, an established baseline is necessary to understand what DT, OT, and IT mean, and what they could look like.

The proposed definitions thoroughly and inclusively fuse views from key sources to provide a basis of clarification for establishing authoritative definitions. The analyses are intended to serve as a catalyst for helping the space community engage in further discussion, alignment, and refinement, facilitating the foundation for an integrated test architecture, specifically the National Space Training and Testing Complex and the operational test and training infrastructure.