Test and Evaluation as a Continuum

Authors: Mr. Christopher Collins and Mr. Kenneth Senechal

Introduction

> Building the future Joint Force that we need to advance the goals of the National Defense Strategy requires broad and deep change in how we produce and manage military capability. (2022 National Defense Strategy of the United States, October 27, 2022, p. 19)

For the United States to maintain an advantage over our potential adversaries, we must make a critical change in how test and evaluation (T&E) supports capability delivery. Making this change requires a new paradigm in which T&E will inform today's complex technology development and fielding decision space *continually* throughout the entire capability life cycle—from the earliest stages of mission engineering (ME) through operations and sustainment (O&S). This transformational shift will significantly strengthen the role of T&E in enabling critical warfighting capability delivery at the "speed of relevance."

The transformation, required in response to our adversaries' aggressive pace of capability development and the unprecedented and accelerating pace of innovation in emerging technologies (artificial intelligence (AI), autonomy, quantum, 6G, etc.), is proposed to address many of the historic challenges that acquisition programs have experienced, while also accounting for the increasing complexity of Department of Defense (DoD) systems. This new approach moves T&E from a serial set of activities conducted largely independently of systems engineering (SE) and ME activities to a new integrative framework focused on a continuum of activities termed *T&E as a Continuum*.[1] This framework has three key attributes: (1) capability and outcome focused testing; (2) an agile, scalable evaluation framework; and (3) enhanced test design. Each attribute is critical in capability delivery at the "speed of relevance."

T&E as a Continuum builds off of the 2018 DoD Digital Engineering (DE) Strategy's five critical goals: (1) formalize the development, integration, and use of models to inform enterprise and program decision-making; (2) provide an enduring, authoritative source of truth; (3) incorporate technological innovation to improve the engineering practice; (4) establish a supporting infrastructure and environments to perform activities, collaborate, and communicate across stakeholders; and (5) transform the culture and workforce to adopt and support DE across the life cycle.[2] Derived from those goals are three key enablers for implementing T&E as a Continuum, developing (1) robust live, virtual, and constructive (LVC)[3] testing environments; (2) a model-based environment; and (3) a "digital" workforce. Each key enabler plays a critical role in conducting T&E as a Continuum, as combined they will provide timely and comprehensive information on system performance and risk characterization from the earliest design stages, enabling rapid development and fielding along with ongoing support for these increasingly complex systems and systems of systems (SoS). Making T&E an integral part of SE and ME processes facilitated by using DE not only supports decision-making on individual

systems but also helps enable informed management of DoD's capability development portfolios.

Background

Traditionally, and especially aligned with Major Capability Acquisition[4] programs, the current T&E paradigm has been shaped and executed as a somewhat serial process, often with discrete test events, proceeding through Contractor Testing (CT), Developmental Testing (DT), and finally Operational Testing (OT), to realize the delivered system capability. This current practice gathers a constrained set of information from a limited number of resource-intensive discrete events. It also does not explicitly account for AI/machine learning (ML)-enabled systems that will learn and change throughout development and into fielding and O&S. Even when adopting some model of Integrated Testing[5] that combines elements of CT, DT, and/or OT, the process remains largely serial and still requires separate evaluation frameworks for DT and OT.[6] CT and DT both stay focused toward requirements verification, whereas OT evaluates mission effectiveness and suitability, generally late in the development cycle. This approach, which defers early consideration and T&E of performance in realistic mission scenarios, has resulted in late "discovery" of issues affecting performance in those scenarios. When manifested, particularly when development begins with immature system technologies and components, this late "discovery" is extremely challenging to address; program planning and funding typically do not incorporate the time and resources needed to comprehensively address significant problems discovered just before fielding. Programs often adjust needed functionality or system performance to deliver on defined timelines, degrading the effectiveness of the capability delivered to the Warfighter.

In addition, the SE community has historically used a serial interpretation of the SE "V"[7] resulting in late engagement of the test community in evaluating system performance in mission-relevant scenarios. The late engagement of the test community delays the critical feedback loop provided by testing, which is required for development and evolution of robust system designs (especially AI/autonomous systems).

Recognizing these challenges, the software engineering community is leading efforts, captured via the Software Acquisition Pathway,[8] to implement a more iterative design/test approach to software development; however, these efforts have yet to be scaled to the level necessary for development and fielding of complex systems incorporating both hardware and software. There are also several Service-led efforts to implement T&E as a Continuum integrated within SE and ME including the Navy's Capabilities-Based T&E (CBTE), which brings a mission task-focused approach to the design and testing of Navy programs.[9]

The DoD T&E enterprise is also developing an overall T&E strategy[10] that responds to the increasingly complex and dynamic threat environment. The Director, Operational Test and Evaluation (DOT&E) is capturing this effort with coordination across all Service T&E

Executives as well as other Office of the Secretary of Defense (OSD) T&E organizations. The strategy, introduced by DOT&E, includes five pillars:

- Test the way we fight.

- Accelerate the delivery of weapons that work.

- Improve survivability of DoD in a contested environment.

- Pioneer T&E of weapon systems built to change over time.

- Foster an agile and enduring T&E enterprise workforce.

These pillars, and the underlying Strategic Implementation Plan,[11] will drive the required changes in the traditional approaches to T&E and will significantly aid in delivering the enablers (discussed later in this article) to implement the proposed T&E as a Continuum concept.

With this background, T&E as a Continuum can be broken down into three key attributes as well as three key enablers (see Figure 1).
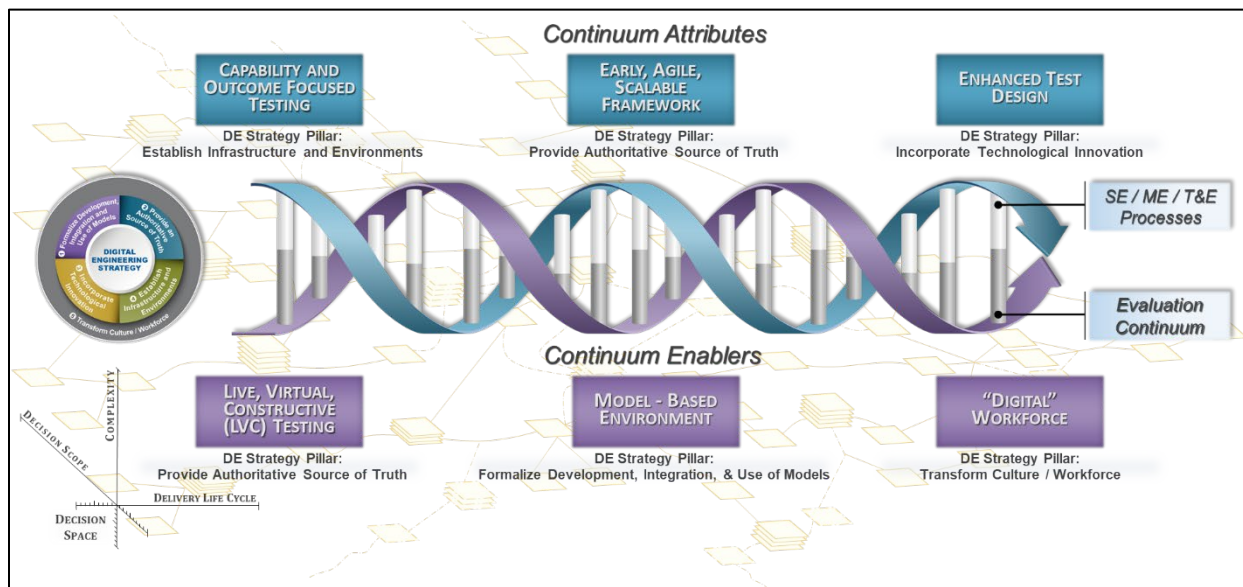

Figure 1. T&E as a Continuum: Attributes and Enablers

Description and Attributes

Executing T&E as a Continuum will combine ME, SE, and T&E into parallel, collaborative, and combined efforts. The collaborative use of mission task decompositions, associated task conditions, and the associated measures of effectiveness and performance produced by ME will allow SE and T&E to develop a common understanding of the design iterations and testing activities needed to validate mission capability. This validated mission capability will be especially critical for systems that learn and continually adapt in the field. This combined approach to conducting ME, SE, and T&E enables early and consistent focus on mission accomplishment rather than requirements verification. Emphasizing the shared responsibilities

between SE and T&E, the continuum enables early SE design trades and analyses to make use of the test methodologies and conditions that OT will ultimately use to validate mission effectiveness. This approach moves T&E from a serial set of activities conducted largely on the right side of the SE "V" to the conduct of a continuum of activities beginning on the left side of the "V" (see Figure 2). The model-based environment and associated tools will allow the SE community to manage the many complex activities being conducted simultaneously within this continuum across the "V." It will enable ME, SE, and T&E communities to work in tandem early and continually and to respond quickly to the dynamic threat environment—ultimately collapsing the "V."
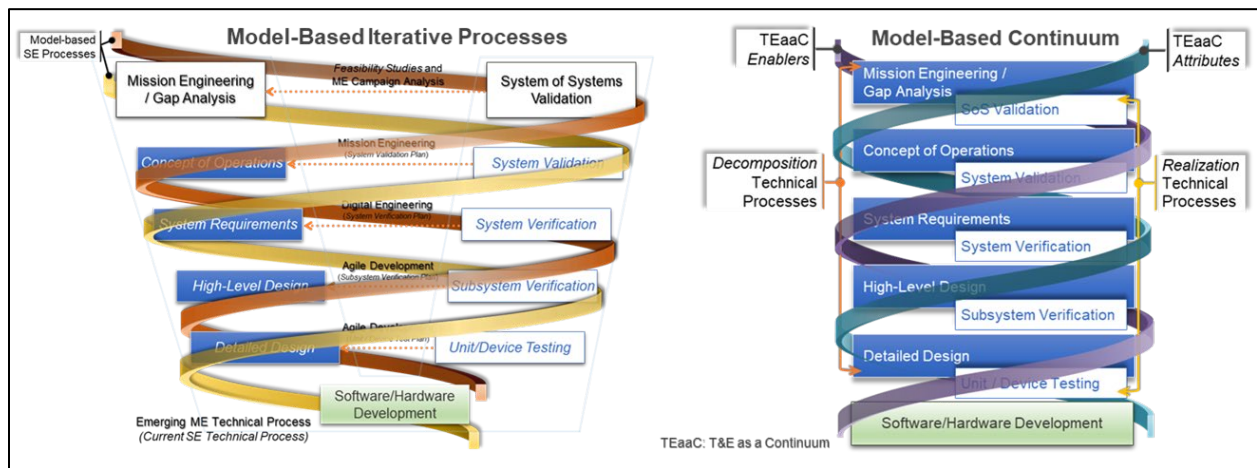

Figure 2. Model-Based Systems Engineering "V" Continuum

The three key attributes for T&E as a Continuum—capability and outcome focused testing; an agile, scalable evaluation framework; and enhanced test design—are described below.

<u>Capability and Outcome Focused Testing</u>. This approach, often called "mission-focused testing," focuses as early as possible on performance of the military capability when fielded. DoD needs to "test as we fight" and ensure that early engineering and technology verification testing, conducted during both CT and DT, provide information that directly supports evaluation of performance in a mission context culminating in OT. This direct alignment of all CT and DT activities to the mission provides the decision maker with a very clear picture of mission-level capability performance throughout the acquisition continuum, as well as earlier insight regarding risks associated with mission capability performance against the current and projected threat environments. This mission context also facilitates the conversation on operational resilience,[12] including, as one example, resilience during operations in highly dynamic or contested environments, including those associated with adverse conditions created by a cyberspace attack.

An early resilience focus for testing will help prioritize responses to testing outcomes in shaping future testing requirements and priorities. In the cyber domain, for example, this focus on resilience starts to move the discussion toward one of mission context, adverse impacts, early detection of those impacts, and better recovery capabilities, significantly expanding from the

"controls" approach to containing impacts of cyber threats.[13] In the future, as the context of "requirements" most likely evolves from discrete Key Performance Parameters and Key System Attributes to one of "system behaviors" (provided via a model of the system), the continuous testing process will allow more rapid evaluation against those desired system behaviors within the context of the model-based environment.[14]

As a critical underpinning, the deliberate execution of the ME process will provide representations of DoD missions and the capabilities needed to conduct them, thereby enabling the identification and update of system requirements. Leveraging a model-based systems engineering (MBSE) approach and using shared capability representations enable iterative T&E of evolving weapon systems, including those incorporating AI/ML.[15]

<u>Agile, Scalable Evaluation Framework</u>. Successfully addressing "decision space" across the capability life cycle from ME through O&S requires a foundational and holistic evaluation framework developed before formal program initiation. This evaluation framework is continually updated, identifying the information needed to support decision-making and how it will be obtained using ME, SE, and T&E activities and tools. Current evaluation framework development efforts begin too late and are not integrated across the Science and Technology (S&T), DT, and OT domains. This new evaluation framework should be adaptable and scalable across the six Adaptive Acquisition Framework pathways,[16] as well as tailorable to pre-Program of Record S&T, prototyping, and experimentation efforts. It will enable iterative and continual T&E refinement as performance information accrues and a program matures. The evaluation framework should include the information needed to perform verification, validation, and accreditation (VV&A) of modeling and simulation (M&S) and other MBSE elements essential to conducting T&E as a Continuum. The evaluation framework should also identify the test infrastructure, including the LVC assets needed to obtain the required evaluation information. This identification will provide an early, clear, and direct demand signal for resourcing T&E infrastructure across DoD's capability portfolios.

A holistic evaluation framework also underpins a consistent "informed risk management" methodology, providing improved understanding of "as is" system performance against the threat environment. This approach is already being applied with current agile software development processes and, coupled with capability and outcome focused testing, will provide decision makers with significantly more insight into alignment or misalignment of design development with threat environments. This approach is important to support early fielding and prototyping decisions, improved understanding of overall system resilience, and improved understanding of the remaining sufficiency and scope of program design and validation efforts required to support the delivery of desired capabilities.

Developing this evaluation framework within a model-based environment will help manage integration complexities across subsystem, system, and especially SoS elements. The framework

will also enable management of verification and validation (V&V) efforts across asynchronous development efforts.

Enhanced Test Design. Developing an integrated test strategy spanning the evolution of specified capabilities over time, while accounting for highly dynamic and contested mission environments, is a critical component of T&E as a Continuum. An enhanced test design identifies the testing needed to provide the information identified in the agile, scalable evaluation framework. Using the evaluation framework as a basis, the enhanced test design will expand upon recent efforts to improve Integrated Testing and better support early evaluation of performance in a mission context. Enhanced test design incorporates VV&A across the *whole* of capability evolution or the program life cycle and helps early engineering and technology verification testing to provide initial data informing whether desired capability outcomes can be achieved. It also incorporates an iterative "design/test/validate" approach comprehensively assessing evolving "system behaviors" and allowing capability development efforts to adapt to the changing operational environment. Enhanced test design should also include robust evaluation of human systems integration focusing on demonstrating the trustworthiness necessary for Warfighter acceptance and utilization of AI/ML-enabled systems. Trust—whether by oversight, procurement, or user communities—depends on this assurance.[17] For the Warfighter, in this context, human systems test design is especially important. In conflict, utilization decisions must often be nearly instantaneous and rely on properly calibrated trust in the system as it is used. Metrics for trust and for teaming will be essential.[18] The enhanced test design also incorporates, as appropriate, Scientific Test and Analysis Techniques[19] to gain efficiencies and help enable rigorous evaluation of test results. Bayesian statistics has recently gained increased attention in this regard, as it offers a mathematically rigorous means to combine predictions and uncertainties from multiple sources and to update them during a series of test activities.[20] Many organizations and Services[21] are already using elements of enhanced test design to develop efficient, comprehensive strategies for conducting T&E.

Enablers

Implementing T&E as a Continuum attributes involves developing three key enablers—robust LVC testing, a model-based environment, and a "digital" workforce—as described below.

Robust LVC Testing. Given the complexity of modern warfighting systems, our concept of conducting T&E has to evolve to be comprehensive across LVC approaches, including model-based testing. Conducting live, open-air testing for some DoD systems presents significant challenges due to system complexity and/or sensitivities to revealing capabilities. Thus, increased use of M&S and other constructive approaches will be essential in obtaining a comprehensive understanding of systems' performance.[22] DoD needs to continue evolving its test ranges and facilities to incorporate a combination of LVC capabilities spanning as much of the threat and operational environment spectrum as feasible. This evolution would entail a "shared" infrastructure within the DE environment (see Figure 3).[23]

Figure 3. Shared Capability Across the Acquisition Life Cycle

One very tangible benefit of this evolution will be the ability to conduct SoS testing, an essential requirement to support evaluation of concepts, such as Joint All-Domain Command and Control (JADC2), in the context of all-domain operations. Given the asynchronous development of the systems comprising SoS capabilities, the LVC framework will significantly improve integration in support of capability validation, especially for distributed or multi-agent AI/ML capabilities of platforms collaborating toward a common goal/objective. These LVC capabilities will also enable our forces to conduct the realistic training necessary to successfully accomplish all-domain operations. The increased emphasis on LVC testing does not diminish the need for live, open-air testing to evaluate key aspects of systems' performance, as well as to support rigorous VV&A of M&S. However, with proper VV&A, LVC testing should enable a more comprehensive understanding of the capabilities that will be provided to conduct multi-domain operations against increasingly sophisticated adversaries.

Model-Based Environment. The model-based environment will be the most critical enabler of T&E as a Continuum. The complexity of our capability development efforts, coupled with the complexity of enhanced test design, will require a digital backbone and model-based approaches to manage the continuum of T&E activities; integrate them with ME and SE activities conducted in parallel; and manage, curate, and analyze the data that all those activities generate. This model-based environment will incorporate the agile, scalable evaluation framework and help ensure its consistency with the full range of ME, SE, and T&E activities. Using this model-based environment, DoD can transition to a "model-test-validate-design-test" process providing

early and continual information on expected mission capability versus the traditional "design-build-test-fix-test" process that too often results in late discovery of capability shortfalls. This "model-test-validate-design-test' process is currently being used.[24] It can include a "digital thread" displaying progress in modeling system performance beginning early in ME through the potential development of high-fidelity "digital twins"[25] to support programs in O&S. To support T&E, the environment must also include realistic representations of the joint battlespace within which systems will operate.

"Digital" Workforce. Underpinning adoption of the model-based environment will be the need for a "digital" workforce that is savvy with the processes and tools associated with MBSE, model-based T&E, and other model-based processes. This workforce transformation has already begun at an initial level with the revamping of the Defense Acquisition Workforce Improvement Act (DAWIA) certification process and associated functional domain credentialing.[26] Additional efforts are necessary to obtain and develop within DoD the technical and management skills needed to lead and manage a continuum of ME, SE, and T&E activities, conducted in parallel using model-based approaches.[27] In addition, current efforts do not drive the full integration required within the broader SE domain to ensure T&E is integral to SE V&V efforts. Lastly, the "collapsing" of the SE "V" does not negate good SE practices; the challenge lies in iteratively managing those practices to ensure alignment with the dynamic, connected nature of this new model-based SE "V" environment.

Challenges and Next Steps

Achieving T&E as a Continuum will require surmounting several challenges:

- As many organizations across DoD and industry have recognized the value of adopting the model-based approaches essential to achieving T&E as a Continuum, they have initiated many critical, but disparate, activities. The T&E community will need a central repository identifying these efforts and the important details regarding their implementation to identify and avoid interoperability issues. This repository will also ensure that multiple different organizations, groups, and programs avoid "reinventing the wheel" by recreating a model, tool, or process already developed by someone else in the DoD enterprise. These efforts should seek to employ a common T&E modeling schema and data dictionaries, aided by coordination among the Services and OSD to develop requisite policy and guidance.

- AI/ML systems that are continually evolving and adapting throughout their life cycle, as well as iteratively developed software systems, will require development and adoption of iterative T&E approaches extending through fielding and O&S to ensure the continued "trustworthiness" of these systems.

- Enabling defensible use of M&S as part of the increased use of LVC capabilities to evaluate system and SoS performance will require adoption of a "Model Validation Level"[28] or similar concept.

- The workforce, along with the SE and T&E functional domains, will need to evolve *and integrate* [emphasis added] their processes to now work in this iterative testing process—in which V&V activities traditionally conducted on the right side of the SE "V" move to the left and are conducted iteratively, recursively, and continually. This change to the traditional SE "V" instantiates a "transdisciplinary"[29] holistic approach for SE and T&E, versus the current "cross-disciplinary" approach that still relies on SE and T&E to apply their own methods to their specific domains.

Achieving T&E as a Continuum will require several key next steps:

- Developing a DoD roadmap, encompassing the described *Attributes* and *Enablers*, for implementing T&E as a Continuum using model-based approaches, focusing initially on new start programs.

- Establishing a core team to identify and assist selected new start capability development efforts, including Programs of Record, with implementation and sharing lessons learned, thereby establishing the basis for subsequent wider adoption across DoD.

- Developing a strategy for consistent reuse of models and data, incorporating model-/data-driven techniques for conducting VV&A, incorporating data standards, and identifying the associated supporting infrastructure.

- Creating a partnership with our industry counterparts to jointly develop key T&E MBSE artifacts including reference models and ontology as well as contracting guidance consistent with conducting T&E as a Continuum environment.

- Developing and implementing across the DoD enterprise an overarching communications strategy explaining the critical need, key elements, and immense value of conducting T&E as a Continuum.

Summary

The overall acquisition paradigm needs to transform for DoD to deliver capability at the "speed of relevance," especially with the accelerating pace of innovation in emerging technologies (autonomy, AI, quantum, 6G, etc.). Rapid implementation of T&E as a Continuum will allow DoD stakeholders to more closely align T&E to support DoD leadership's "decision space." T&E as a Continuum will provide "informed risk management," allowing earlier understanding of progress in development and confidence across the acquisition life cycle that required capabilities will be delivered when needed to support the Warfighter. This new approach is critical to ensure DoD's ability to rapidly respond to changes in the current and future operating environment, including rapid advances in threat capabilities and concepts of operations and employment. Improvements in risk characterization and overall assurance are critical given the complexity of current DoD capabilities, and T&E as a Continuum achieves those improvements. These efforts will naturally drive an integration and intertwining of our traditional SE and T&E activities, providing a transdisciplinary, continuous process coupling design evolution with

VV&A. Conducting T&E as a Continuum will enable more responsive capability delivery to meet the Warfighter's needs.

Mr. Christopher Collins is the Executive Director, Developmental Test, Evaluation, and Assessments in the Office of the Under Secretary of Defense for Research and Engineering. Mr. Kenneth Senechal is the Capabilities-Based T&E Director in the Office of the Commander, Naval Air Systems Command.

Significant Contributors: Mr. Todd Standard (Naval Air Warfare Center Weapons Division), Dr. Michael Gilmore (Institute for Defense Analyses), and Mr. Hans Miller (MITRE).

Notes

[1] The Cambridge Dictionary defines continuum as "something that changes in character gradually or in very slight stages without any clear dividing points," https://dictionary.cambridge.org/us/dictionary/english/continuum. The Merriam-Webster Dictionary defines continuum as "a coherent whole characterized as a collection, sequence, or progression of values or elements varying by minute degrees," https://www.merriam-webster.com/dictionary/continuum.

[2] Office of the Deputy Assistant Secretary of Defense for Systems Engineering, "Department of Defense Digital Engineering Strategy," June 2018.

[3] LVC is the categorization of simulation into live, virtual, and constructive classes or divisions; together LVC indicates the integration of all three into a particular event or activity. (DoD Modeling and Simulation Body of Knowledge)

[4] DoD Instruction 5000.02, "Operation of the Adaptive Acquisition Framework," January 23, 2020, as amended.

[5] DoD Instruction 5000.89, "Test and Evaluation," November 19, 2020, pp. 7–10. "Integrated testing requires the collaborative planning and execution of test phases and events to provide shared data in support of independent analysis, evaluation, and reporting by all stakeholders. Whenever feasible, the programs will conduct testing in an integrated fashion to permit all stakeholders to use data in support of their respective functions."

[6] DoD Instruction 5000.89, p. 9. The Integrated Decision Support Key (IDSK) is defined as "A table that identifies DT, OT, and LF [live fire] data requirements needed to inform critical acquisition and engineering decisions (e.g., milestone decisions, key integration points, and technical readiness decisions). OT&E and DT&E will use the IDSK to independently develop evaluation frameworks or strategies that will show the correlation and mapping between evaluation focus areas, critical decision points, and specific data requirements." Unfortunately, the current IDSK approach still requires independent DT and OT evaluation frameworks that exacerbate the "serial" thinking through the T&E life cycle.

[7] Adapted in the International Council on Systems Engineering (INCOSE) Systems Engineering Body of Knowledge from Kevin Forsberg, Hal Mooz, and Howard Cotterman. However, the V-Model's history is unclear. It appears to have its roots in the 1960s, though public citations are scarce. It likely emerged from several sources somewhat independently. The first reference in a systems engineering context seems to be at the first annual conference of INCOSE in 1991, the predecessor to the INCOSE International Symposium. Kevin Forsberg and Hal Mooz spoke to the V-Model in terms of iteration and increments and included a three-dimensional version. The three-dimensionality recognized a single top level of systems context and a single realized integrated system to meet business or mission needs, while the lower levels partitioned the system iteratively, with each partition having its own V-Model. That is, a system has subsystems that each have their own V-Model, each subsystem in turn partitions to components with their own V-Models, and so forth. The V-Model is intended to be read vertically in terms of the multiple levels of a system interest, with the top level referring to system context and the lowest level the parts or elements of the system. The horizontal interpretation of the V-Model denotes a logical sequence of the value stream. Not all requirements need to be imposed a priori—an increment in development may permit developing some system elements while continuing to define requirements that remain at the top level. (Adapted from Barnum, Winstead, Fisher, Walsh, "A Classical Modernization of the V-Model," 2022 (DRAFT))

[8] DoD Instruction 5000.02, "Operation of the Adaptive Acquisition Framework," January 23, 2020, as amended, p. 13.

[9] CBTE is captured in SECNAVINST 5000.2G and is being implemented by Office of the Chief of Naval Research and Director of Innovation, Technology Requirements, and Test and Evaluation/Department of the Navy T&E (OPNAV N94/DoN T&E). CBTE leverages a partnership between DT and OT to collaboratively execute a mission-based task decomposition and test design pre-Milestone A/B that aligns all CT, DT, and OT to mission accomplishment. The test framework generated by this process informs early infrastructure decisions along with providing mission-informed reporting throughout test execution by directly linking test results to mission accomplishment.

[10] Office of the Director, Operational Test and Evaluation, "DOT&E Strategy Update 2022," June 13, 2022. Additional details on each of the five strategic pillars can be found in the document.

[11] Office of the Director, Operational Test and Evaluation, "DOT&E Strategy Update 2022 Implementation Plan," November 1, 2022 (DRAFT).

[12] Operational resilience: The ability of systems to resist, absorb, and recover from or adapt to an adverse occurrence during operation that may cause harm, destruction, or loss of ability to perform mission-related functions. (DoD Instruction 8500.01, "Cybersecurity," March 14, 2014, as amended)

[13] DoD currently implements mission-based cyber risk assessments as required by DoD Instruction 5000.89 and explained further in the current DoD Cybersecurity T&E Guidebook (Version 2.0, Change 1, February 10, 2020). Updated details will be published in 2023 in the forthcoming DoD Manual on Cyber T&E and the DoD Cyber T&E Companion Guide (the replacement to the DoD Cybersecurity T&E Guidebook).

[14] Model-based engineering enables tracking of complex relationships between system components and the capabilities they implement in a mission context. By creating executable descriptive models that tie into M&S, testing against simulation can become continuous, as in continuous integration/continuous delivery software development. The challenge will then fall upon the T&E community to perform agile VV&A to ensure the models are correct.

[15] T&E as a Continuum is intended to inform and support ongoing DoD efforts to revitalize Capability Portfolio Management (CPM). The Office of the Under Secretary of Defense for Acquisition and Sustainment is updating DoD Directive 7045.20, "Capability Portfolio Management." The three key aspects of CPM in this update include the following: (1) Drive strategic alignment across planning, requirements, technology, acquisition, budgeting, and execution. Senior leadership will use CPM as part of the DoD's decision support systems, to inform capability improvements through the lens of joint, integrated mission effects; (2) Be data driven. CPM will leverage schedule, cost, and performance metrics for systems and interdependencies among portfolio elements to identify options to close capability gaps and synchronize development and fielding priorities; and (3) Be focused on maximizing capability effectiveness, enabled through mission integration management and mission engineering, to advise DoD leadership on ways to optimize capability portfolios and ensure that they are strategy driven, achievable, and affordable and balance near-term and long-term objectives.

[16] DoD Instruction 5000.02, "Operation of the Adaptive Acquisition Framework," January 23, 2020, as amended.

[17] David M. Tate, "Trust, Trustworthiness, and Assurance of AI and Autonomy," IDA Document D-22631, April 2021.

[18] Heather M. Wojton, Daniel Porter, Stephanie T. Lane, Chad Bieber & Poornima Madhavan. "Initial validation of the trust of automated systems test (TOAST)." *The Journal of Social Psychology* 160(6):735–750, 2020. DOI: 10.1080/00224545.2020.1749020

[19] Scientific Test and Analysis Techniques (STAT) are the scientific and statistical methods and processes used to develop efficient, rigorous test strategies that yield defensible results. Examples include design of experiments, model validation, regression analysis, survey design, uncertainty quantification, and

Bayesian statistics, each used within a larger decision support framework. The suitability of each method is determined by the specific objective(s) of the test, delivering insight to inform better decisions. (Multiple authors, "Guide to Developing an Effective STAT Test Strategy V8.0," STAT Center of Excellence, December 31, 2020, https://www.afit.edu/STAT/passthru.cfm?statfile=59)

[20] In Bayesian statistics, the prediction and uncertainty of a model parameter are represented by a probability distribution. This distribution is continually updated as new information is obtained through modeling and simulation, live testing, or other sources. As more data become available, this distribution ideally becomes narrower and shifts toward the new evidence, indicating enhanced understanding of the system as represented by the model. Bayesian statistics can be a valuable tool, but implementing it requires expertise. (Maj Victoria Sieck, PhD, and Dr. Kyle Kolsti, "Bayesian Methods in Test and Evaluation: A Decision Maker's Perspective," July 2022, https://www.afit.edu/STAT/passthru.cfm?statfile=148)

[21] For example, see United States Space Force, "Space Test Enterprise Vision," March 2022, p. 2; "The USSF test philosophy aims to bridge the historical stovepipes between developmental test (DT) and operational test (OT) in order to infuse operational perspective early and speed delivery of credible, war-winning capabilities in the face of a rapidly evolving threat."

[22] The synthetic data generated will be essential in accelerating AI/ML capabilities.

[23] The Test Resource Management Center (TRMC) has several ongoing initiatives addressing development of this shared infrastructure and can provide additional information as requested.

[24] For example, see "Accelerating Innovative Digital Engineering Capabilities to the Warfighter," Breaking Defense, September 29, 2021, available at https://breakingdefense.com/2021/09/accelerating-innovative-digital-engineering-capabilities-to-the-warfighter/, accessed December 12, 2022.

[25] A digital twin is the electronic representation—the digital representation—of a real-world entity, concept, or notion, either physical or perceived. The defining feature of a digital twin is the ongoing data integration between the digital model and its physical unit counterpart. Digital models used for research and development can evolve into digital twins of specific units once they are produced and fielded, and these twins can continue to be updated as their underlying digital model is refined. (NISTIR 8356 (Draft) modified by Director, Operational Test and Evaluation, "Digital Twin Assessment, Agile Verification Processes, and Virtualization Technology," July 2022) A digital thread is a data-driven architecture that links together information generated from across the product lifecycle and is envisioned to be the primary or authoritative data and communication platform for [an organization's] products at any instance of time. (Victor Singh and Karen E. Willcox, "Engineering Design with Digital Thread," 2018) Digital threads give engineers access to a complete definition of a system and the mission context from which it was derived.

[26] Tom Simms, "DoD Acquisition Workforce T&E Functional Area," Functional Integration Team Meeting briefing, June 2022. The DoD T&E Functional Integration Team is leading efforts for the T&E community regarding changes to the overall DAWIA certification process. DAU recently established two levels of T&E certification, Foundational and Practitioner. In addition, there are a number of T&E credentials in development addressing focused topics within the T&E community. These credentials will be aligned with specific billet tasks and responsibilities, and will be time-limited, requiring recertification as appropriate.

[27] Potential opportunities include new hiring authorities, new Defense Acquisition University courses, and opportunities to conduct coursework at top U.S. engineering schools.

[28] A Model Validation Level (MVL) is a single objective metric for how much trust can be placed in the results of a model to represent the real world. The calculation is based on measures of fidelity (model-to-referent prediction comparisons), referent authority, and scope (intended use space versus data coverage).

The result is a real number from zero (0) to nine (9), which can be no greater than the authority value transferred from the referent. The MVL enables an objective, continually updatable approach to model validation that captures validity in the intended use conditions as a score rather than a binary "pass/fail" judgment. (Kyle Provost, Corinne Weeks, Nicholas Jones, and Maj Victoria Sieck, PhD, "Elements of a Mathematical Framework for Model Validation Levels," Scientific Test and Analysis Techniques Center of Excellence, August 2022, retrieved from https://www.afit.edu/STAT/statdocs.cfm?page=1126)

[29] Systems Engineering is a transdisciplinary and integrative approach to enable the successful realization, use, and retirement of engineered systems, using systems principles and concepts, and scientific, technological, and management methods. (International Council of Systems Engineering (INCOSE), https://www.incose.org/about-systems-engineering/system-and-se-definition/systems-engineering-definition, accessed December 23, 2022)