After DOT&E:
Reforming Test and Evaluation for the Age of Lethality

Dr. Marcus Jones
July 29, 2025

ETC
Energetics Technology Center

Executive Summary

This think-piece examines the implications and potential of the May 2025 directive reorganizing the Office of the Director of Operational Test and Evaluation (DOT&E), a reform aimed at increasing agility, reducing bureaucratic friction, and focusing the Department of Defense's test and evaluation (T&E) enterprise on its core statutory mission. The reorganization marks a turning point in the evolution of oversight and performance assessment for defense systems, one that invites fresh thinking about how best to align speed, innovation, and warfighter confidence. The urgency of this reform has now been explicitly acknowledged in Congress: the Senate's FY26 NDAA includes a legislative proposal to establish an Alternative Test and Evaluation Pathway, initially scoped to software-intensive systems, that embodies many of the very principles advocated here: mission-focused evaluation, continuous feedback, early failure discovery, and decoupling from rigid documentation requirements.

Drawing on four decades of institutional experience, this paper explores the rationale for reimagining T&E as an integrated, continuous function grounded in mission context, powered by digital tools, and focused on fielding capabilities that are both effective and adaptable. It highlights how legacy structures, while built on good intentions, have often struggled to keep pace with the demands of software-defined systems, autonomous platforms, and modern joint operations.

The paper identifies key enablers that can help ensure the success of the current transformation: investment in digital test infrastructure, reinforcement of evaluation as a lifecycle function, preservation of transparent performance reporting, and the development of a modern T&E workforce. These steps are not about preserving legacy forms but about building a leaner, faster, and smarter T&E system aligned with emerging technologies and operational demands.

_____

On May 27, 2025, the Secretary of Defense issued a directive mandating a comprehensive reorganization of the Office of the Director of Operational Test and Evaluation (DOT&E).[1] Framed explicitly as a strategic move to enhance the lethality, readiness, and efficiency of the U.S. Armed Forces under an "America First" defense strategy, the directive ordered the most sweeping structural reduction of the office since its establishment in 1983. Its intent was to return the office to what was described as a "zero-based core mission," sharply narrowing its

_____

[1] 27 May 2025, "Reorganization of the Office of the Director of Operational Test and Evaluation," Memorandum for Senior Pentagon Leadership; see also Ashley Roque, "Hegseth reshaping Pentagon's weapons testing oversight office, cutting staff positions," *Breaking Defense*, May 28, 2025.

focus to the fulfillment of statutory responsibilities under Title 10 of the U.S. Code, and shedding a host of functions deemed non-essential, non-statutory, or duplicative. DOT&E's civilian staff is to be reduced from 118 to just 30 positions, with a single Senior Executive Service member and no more than 15 assigned military personnel. The memorandum also calls for a full reduction-in-force, placing all non-retained personnel on administrative leave within seven days of the directive. Contractor support to the office is to be terminated within the period, and any future contracting activity must be justified anew, subject to Deputy Secretary review and only after a 60-day acclimation period. In parallel, leadership of the reorganized DOT&E has been entrusted to Carroll Quade, formerly the Navy's Deputy for Test and Evaluation, who must now oversee the transition while maintaining compliance with statutory oversight requirements. While some critics have viewed this move as a gutting of independent test oversight, it may equally be read as clearing institutional space for new legislative and structural mechanisms that preserve oversight while enabling agility, most notably the Alternative T&E Pathway now proposed by the Senate in the FY26 NDAA.

Although framed in administrative and budgetary terms, the directive also reflects broader concerns about the role and responsiveness of centralized oversight. Critics have long argued that DOT&E's processes, while intended to ensure rigor, have at times introduced friction into the development and fielding of advanced systems, particularly where timelines are compressed or requirements are evolving. In this view, the office's expansive interpretation of its mandate has occasionally imposed coordination and validation burdens that exceeded statutory requirements, slowing programs that aim to move at the pace of operational need. The new directive appears to reflect that perception. It effectively recasts DOT&E's role from an expansive oversight body with broad evaluative authority to a lean statutory compliance node. With the elimination of contractor support, the curtailment of leadership positions, and the removal of non-statutory activities, the office will no longer perform many of the integrative, cross-service, and anticipatory functions that had evolved over four decades of institutional practice. What remains is a narrow focus on certifying that designated major defense acquisition programs meet the legal thresholds for operational test and live fire test reporting, with minimal additional infrastructure or analytical capacity. The directive represents, in other words, not a termination of the essential mission of operational T&E, but a redefinition of how it will be fulfilled and by whom.

Since its establishment, DOT&E's evolution has reflected a continuous balancing act between independence and integration, statutory rigor and programmatic flexibility, and legacy practices and emerging needs. But its mission has remained constant: namely, ensuring that weapons work in combat. Of course, *how* it carries out that mission have shifted in response to changing technologies, acquisition policies, operational demands, and political oversight. DOT&E was established by statute in 1983 during a period of growing concern within Congress over the

reliability, suitability, and survivability of major defense acquisition programs. Spearheaded by the Congressional Military Reform Caucus, the legislation arose from growing bipartisan frustration with service-led acquisition programs that bypassed rigorous operational evaluation. The Sergeant York air defense system, which failed spectacularly in live testing and was canceled shortly after DOT&E deemed it ineffective, became the office's first defining case[2] and showed that rigorous testing must be the essential gatekeeper, discerning workable solutions from flawed ones before sunk cost, contractual inertia, or political pressure force suboptimal systems into the field. Even more notorious was the controversy surrounding the Bradley Fighting Vehicle, where efforts by Colonel James Burton to enforce realistic live-fire testing revealed fatal vulnerabilities previously concealed by sanitized test procedures. These and similar episodes underscored the need for an independent authority to ensure that weapons were tested under combat-realistic conditions before being fielded, leading Congress to mandate DOT&E's creation as a structurally independent body reporting directly to the Secretary of Defense.

As systems grew in complexity, Congress created DOT&E as an independent office within the Office of the Secretary of Defense, charged with ensuring that operational test and evaluation (OT&E) and live fire test and evaluation (LFT&E) were conducted under conditions representative of actual combat. From the outset, DOT&E's role was not simply technical but institutional: it was conceived as a structural counterweight within the acquisition system, funded through its own Program Element in the Defense-wide budget and empowered to report directly and independently to both the Secretary of Defense and Congress.[3] Title 10 of the U.S. Code (sections 139, 4171, and 4231) defines the office's legal authorities, requiring that major acquisition programs undergo OT&E before proceeding to full-rate production, and mandating that DOT&E provide an independent assessment of a system's operational effectiveness, suitability, and survivability. Over time, these core responsibilities were supplemented by oversight of a range of related initiatives, each aimed at improving the quality, realism, and analytic rigor of defense testing across platforms and domains.[4]

---

[2] Sergeant York Division Air Defense System (DIVAD) was a self-propelled anti-aircraft gun that had been rushed into production based on inadequate developmental testing. During operational testing, it famously mistook the exhaust fans on a latrine building for enemy aircraft and fired at them. See Ditton, M.H., 1988. "The DIVAD Procurement: A Weapon System Case Study." *Army Law.*, p.3.

[3] As Dan Grazier recounts, Congress established DOT&E in 1983 specifically because Service-led testing had produced overly optimistic reports and missed serious flaws, and to ensure weapons are proven effective and safe under realistic combat conditions. The office's "integrity and independence," as he puts it, have long been seen as obstacles by some in DoD who prefer less transparency when tests reveal problems: see Dan Grazier, "Are We in a Pentagon Wars 2.0?" Stimson Center, 2025.

[4] These include the Joint Test and Evaluation Program (JT&E), the Joint Technical Coordinating Group for Munitions Effectiveness (JTCG/ME), the Joint Aircraft Survivability Program (JASP), the Center for Countermeasures (CCM), among others.

In the 1990s and 2000s the office grew in both scope and staff as it responded to an expanding acquisition portfolio and heightened congressional scrutiny of cost overruns and performance shortfalls. It developed capacity not only to evaluate test plans and observe trials, but to influence the actual methodological design of test campaigns, the integration of modeling and simulation, and even the independent validation of operational environments. Its role became particularly salient at the height of the Global War on Terror, when systems were being fielded rapidly in response to urgent operational needs. The office had to navigate the tension between acquisition speed and evaluative integrity, balancing congressional insistence on rigorous oversight with the warfighter's demand for timely delivery. In more recent years, DOT&E has emphasized transformation through digital modernization and mission-driven innovation. Its 2023 *Annual Report* highlighted initiatives to integrate high-fidelity modeling and simulation, develop digital threat environments, and increase test realism through constructs like the Joint Simulation Environment (JSE).[5] The report underscored the importance of assessments that integrate data from across the test lifecycle – including contractor testing, developmental testing, operational testing, and modeling and simulation – to support decision-making in dynamic, multi-domain operational contexts.

The office's development has generally tracked with broader shifts in defense acquisition policy. For instance, post-Goldwater-Nichols reforms in the late 1980s and 1990s solidified DOT&E's position in the defense bureaucracy, while the increased emphasis on rapid prototyping, software-defined systems, and Middle Tier Acquisition in the late 2010s and early 2020s introduced new pressures on its traditional oversight role (which have arguably never gone away). DOT&E has adapted over the years by developing guidance for evaluating AI-enabled and autonomous systems, updating its T&E master planning processes, and implementing practices for the verification, validation, and accreditation (VV&A) of modeling tools across the acquisition lifecycle. In fulfilling these roles, it is undeniable that DOT&E evolved beyond its statutory baseline. It gradually acquired responsibility for broader integration of modeling and simulation, led methodological improvements in test realism, and administered key elements of the T&E ecosystem such as the Joint Test & Evaluation Program, the Joint Munitions Effectiveness Manual, and various survivability and countermeasures programs. At its peak, it oversaw testing for over 260 major programs and administered a budget of nearly $400 million. Its staff – a blend of career civilians, military assignees, and contractors, included specialists in weapons effects – modeling and simulation, human systems integration, and cyber survivability.

---

[5] Annual Reports are helpfully archived by the office here; see also Guertin, H.N. and Goodman, A., "Operational Test and Evaluation (OT&E) for Rapid Software Development." *Journal of Test and Evaluation*, vol. 44, issue 1 (March 2023).

Within the broader $150+ billion RDT&E enterprise, DOT&E's funding and authority were comparatively modest but uniquely positioned for enterprise-wide test oversight.[6]

The performance of the office has long been the subject of debate, especially as the strategic and technological landscape has evolved. Program offices sometimes describe the DOT&E process as a 'black box' – a sentiment echoed by developers and engineers who describe long periods of uncertainty, opaque feedback cycles, and post hoc demands for new tests unanticipated in initial requirements. In some cases, these demands introduced critical insights. In others, they yielded significant rework, without clarity about the evidentiary standards being applied. While many in the defense oversight and evaluation community rightly regard DOT&E as a critical safeguard against the fielding of unproven or underperforming systems, acquisition professionals and program managers are not shy about expressing frustration at what they see as delays, duplication, and a misalignment between testing practices and the realities of rapid technological change. The Senate proposal addresses precisely this tension: maintaining independent oversight while eliminating pre-approval veto authority, mandated formats, and milestone-driven gates, while affirming the shift toward lifecycle evaluation grounded in raw data. It requires government ownership of test data, open access for DOT&E, and mandates continuous data collection and learning throughout system development.

These structural tensions have long been especially pronounced in the field of munitions and energetics. As explained in the National Energetics Plan (released by OUSD R&E in 2023) and the Energetics Technology Center's "Energetics and Lethality" study (2021), a series of persistent and mutually reinforcing institutional challenges has undermined the transition of new and more effective energetic materials into operational systems. Among the most prominent are antiquated test and evaluation standards and infrastructure, which have not kept pace with the emergence of novel energetic effects (e.g., multiphase blast), faster munitions, and longer-range systems, whether in evaluating offensive lethality or assessing the survivability and vulnerability of platforms subjected to these advanced threats. Many of these new designs exceed the technical and diagnostic capabilities of Cold War-era T&E ranges and instrumentation, which has resulted almost inevitably in an under-characterization of effects and a lack of empirical support for modeling and simulation validation. And while the independence of the office has served as a necessary check on untested assumptions, it has at times led to test requirement changes late in a program's development cycle, creating schedule and cost pressures, particularly when earlier coordination or integration was lacking.

These criticisms of the T&E system have not emerged in isolation. Rather, they reflect growing recognition across government and industry that the legacy approach to testing – defined by

---

[6] Congressional Research Service, *Defense Primer: Research, Development, Test, and Evaluation* (Nov. 19, 2024), IF10553.

Service stovepipes, lengthy qualification protocols, and high-cost, low-frequency range events – cannot match the tempo or complexity of contemporary acquisition. Indeed, as early as 2007, the '231 Report' and subsequent DoD policy shifts sought to re-balance test and evaluation around integrated, evaluation-focused life-cycle approaches, a vision only partially realized.[7] The acquisition system's reliance on predictive performance claims, formalized in milestone decisions and sustained by requirements documents, has repeatedly outpaced empirical demonstration. In program after program, paper promises have stood in for operational realism, with minimal testing used to justify major production decisions. This misalignment between confidence and evidence lies at the heart of the test and evaluation crisis. Reforms must begin by inverting this logic: demonstration must precede commitment, and real-world performance, not speculative metrics, must drive investment. Particularly in a strategic environment where peer competitors are fielding improved weapons systems with rapid iteration cycles and enhanced performance, the U.S. munitions enterprise must find ways to validate lethality more quickly and comprehensively.[8] *To be clear: the concern is not that the technical practice of test and evaluation lacks rigor or sophistication. Rather, the issue is that the institutional framework within which T&E has been executed, anchored by DOT&E as the central overseer, has proven difficult to adapt.* Even as advanced modeling techniques, high-performance computing, and synthetic environments have become available, the shift to an integrated, model-informed, and data-driven T&E enterprise has been slow.[9] The result is a system that can be technically sound but procedurally inflexible, unable to accommodate the tight timelines and risk tolerance required by next-generation munitions programs.

Thus, the case for seeing the new directive as an opportunity arises not from the notion that DOT&E has failed in its statutory mission, but from the growing gap between what that mission

---

[7] Mosser-Kerner, Darlene. "Test and Evaluation Policies and Practices: a new emphasis." *ITEA Journal of Test and Evaluation*, 29 (2008) pp.157-159; also Beers, S.M., Hutchison, S. and Mosser-Kerner, D., 2013. "Developmental test evaluation framework: Focus on evaluation and analysis for acquisition success." *Phalanx*, 46(3), pp.36-39.

[8] In an August 2024 study, the Defense Science Board argued that today's test processes must evolve to keep pace with technology and the need for speed, and recommended a "strategic shift" beyond the traditional sequential acquisition T&E framework, including continuous testing, better integration of developmental and operational test phases, and earlier involvement of new technologies in realistic trials. The report's top recommendations call for using digital engineering and modeling & simulation to automate and accelerate testing, upgrading test infrastructure (ranges, labs, instrumentation), and investing in the T&E workforce to enable faster, more agile testing without sacrificing rigor: Defense Science Board, "Protecting the Warfighter: Tailoring Test and Evaluation for the 21st Century," August 2024.

[9] While slightly older, a 2015 GAO study rebuts arguments that testing oversight unduly delays programs. After reviewing dozens of acquisition programs, it found that disagreements between DOT&E and program offices were rare (under 10% of programs from 2010–2014) and usually resolved with only modest schedule/cost impacts. In the few cases (e.g. F-35, the CVN-78 carrier), GAO concluded DOT&E's concerns were valid and often led to improvements (for instance, more realistic shock tests on the *Ford* carrier, pushed by DOT&E, were ultimately mandated by Congress). Its findings strengthen any argument that cutting DOT&E in the name of "speed" is a false economy: U.S. Government Accountability Office, "DOD Operational Testing: Oversight Has Resulted in Few Significant Disputes and Limited Program Cost Increases," GAO-15-503 (July 2015).

requires in a modern context and how the office has been structured to fulfill it. The directive of May 2025 can be read as a formal acknowledgment of this gap: a decision to reduce centralized burden, devolve responsibility for test design and integration to the Services and program offices, and create space for a new T&E paradigm better suited to the technical characteristics and acquisition timelines of advanced weapons systems. The 2022 DOT&E Strategy Update and its successors make explicit the imperative to "keep pace with rapid evolutions in technologies, threats, and operating environments," calling for transformation across multiple fronts, including operationally relevant testing earlier in the development lifecycle, expanded use of advanced modeling and simulation and high-fidelity digital twins, and automation of test workflows to shorten cycle times. Far from an external imposition, these reforms echo priorities voiced by the T&E community itself, which increasingly recognizes the value of embedding test and evaluation into early R&D, integrating developmental and operational testing, and leveraging big data and AI tools for test design and analysis, while working toward the more consequential goal of aligning and coordinating testing across phases so that results, wherever generated, can be broadly accepted and inform credible operational evaluation.[10] The vision now emerging across government, industry, and academia is one in which digital testbeds, joint modeling environments, and continuous evaluation replace legacy, serial-stage paradigms. In this light, the May 2025 directive reflects not the retrenchment of DOT&E's mission, but its adaptation to ensure relevance and credibility under modern acquisition and technological conditions.

If the directive signaled a structural reset, the Senate's FY26 NDAA proposal offers a policy architecture to match. The proposed Alternative Test and Evaluation Pathway creates a new option for designated programs to pursue continuous, mission-based evaluation, aligned with a unified DT/OT strategy and free of rigid milestone formats or pre-approved TEMP requirements. DOT&E's role under the pathway is redefined: no longer a gatekeeping reviewer of plans and formats, but a continuous assessor of data sufficiency, learning rates, and risk to mission effectiveness. In this, the proposal operationalizes much of what this paper recommends: shifting from procedural compliance to data-driven validation; empowering services while preserving independent review; and aligning T&E timelines with the realities of digital, software-rich systems. Though initially scoped to the Software Acquisition Pathway, the legislation provides a concrete test case for broader application. It will succeed only if its spirit – manifested in empirical rigor, operational realism, and unfiltered feedback – remains intact.

Whether the precise method of dramatically restructuring DOT&E is the optimal one remains an open question. But what is clear is that the legacy structure had become increasingly misaligned

---

[10] Heather J. Patrick et al., "Positioning Test and Evaluation for the Digital Paradigm," *ITEA Journal of Test and Evaluation* 44, no. 2 (2023) pp.160–180; the article's authors include the then-Deputy Director and -Chief Scientist of DOT&E.

with both the strategic environment and the requirements of an acquisition system that must innovate rapidly and field lethality at scale. However, even as the directive eliminates or devolves many of DOT&E's functions, it does not eliminate the enduring operational requirement for rigorous, independent test and evaluation. That need remains, and the question now becomes how best to meet it under the paradigm created by the directive, the most immediate consequence of which is a decentralization of responsibilities. By eliminating non-statutory functions and dramatically reducing DOT&E's footprint, it effectively pushes the obligation for test design, execution, and assessment down to the Services, program executive offices (PEOs), and program managers. In practice, this means that the oversight and integration tasks previously performed by a central office – ensuring consistency in test methodology, evaluating mission effectiveness and lethality across domains, and validating modeling and simulation tools – must be assumed by organizations that are often resource-constrained and focused narrowly on specific programs.

For weaponeering and T&E engineers, the concern is not abstract. Many worry that in the absence of a unifying architecture or gatekeeping mechanism, the enterprise may fragment, with individual Services and programs developing bespoke test regimes that may be inconsistent, duplicative, or insufficiently rigorous.[11] This is especially problematic in the domain of energetics, where the characterization of novel effects, validation of simulations, and safety certification involve complex technical processes and require specialized infrastructure. Without a shared set of standards and sufficient cross-service coordination, the risks of under-testing, mischaracterization, or failure to identify critical design flaws increase substantially. Such risks are compounded when testing methods remain tailored to legacy materials or platforms, rather than emerging ones. As one very experienced chemical engineer has noted, testing a high-performance energetic like CL-20 using criteria developed for HMX can yield misleading conclusions, penalizing innovation through inappropriate baselines. Tailoring is necessary; fragmentation is not.

But for the munitions and weapons acquisition enterprise broadly, the directive presents both an inflection point and a range of emerging opportunities. Reforming test and evaluation cannot succeed in isolation from broader acquisition reform. Legacy budget systems – structured around long, inflexible program baselines and annual line-item appropriations – discourage adaptive testing strategies and prohibit funding for exploratory evaluation beyond narrowly defined contractual milestones. Real T&E reform depends on a deeper shift toward portfolio

---

[11] John A. Tirpak, "Cut to DoD's Test Enterprise Alarms Critics, Fearing Loss of 'Honest Broker'," *Air & Space Forces Magazine,* June 2, 2025; the point is made even more strongly by Dan Grazier, who argues that any upfront savings "will eventually be drastically eclipsed by the money wasted fielding faulty weapons" and that with an 80% manpower cut, DOT&E will assess test sufficiency based on data collected and shared by Services, as mandated under the alternative pathway, while retaining full access and evaluation authority; see Dan Grazier, "Gutting military testing office may be the deadliest move yet," *Responsible Statecraft,* June 4, 2025.

management, modular development, and government re-assumption of the technical baseline, shifts already articulated by leading voices in defense acquisition. By shifting evaluative responsibility closer to the Services and program offices, the directive may significantly reduce timelines for fielding weapons systems incorporating novel warhead technologies, new propellants, or alternative synthesis pathways for energetic materials and other new features. It could enable greater flexibility in aligning test and evaluation (T&E) practices with programmatic needs, particularly in programs seeking rapid integration of next-generation munitions into platforms across the Joint Force.

At the same time, the withdrawal of centralized, independent oversight introduces new complexities and potential vulnerabilities. Historically, DOT&E's structural independence enabled it to serve as a checkpoint for empirical validation across Service boundaries, offering Congress and senior defense leadership an unfiltered view of a system's operational effectiveness, suitability, and survivability. The office's role in managing modeling and simulation validation, cross-domain test coordination, and lethality metric standardization has been especially significant for systems relying on complex, high-risk performance requirements. Without these functions resident in DOT&E, the burden of empirical integrity now shifts to program managers and Service test organizations, which may vary in their approaches and risk tolerances. Indeed, one of the clearest risks in devolving oversight is the potential erosion of confidence in weapon performance under combat conditions, underscoring the value of independent, operationally informed validation. The integration of advanced capabilities – such as collaborative seekers, multi-mode target recognition, or inflight retasking – complicates the performance envelope and heightens the need for mission-representative testing. In such contexts, the assurance model cannot rely solely on internal program confidence or contractor assertion. It must include testing architectures capable of confirming performance under degraded communications, contested electromagnetic spectrum, and realistic adversary behavior. The former DOT&E function helped institutionalize such realism; the absence of a successor mechanism raises the prospect of premature fielding and brittle systems under operational stress.

The reform, therefore, must be seen not as a diminution of the test and evaluation function, but as a redistribution of where and how that function is performed. The central oversight role may be contracting, but the underlying need and demand for rigorous validation remains, particularly in a strategic environment marked by contested lethality, peer adversary pacing, and the imperative of technological overmatch. *In this light, the directive may be read as an opportunity to construct a more agile, integrated, and forward-leaning T&E enterprise, one in which early-phase, platform-independent testing infrastructure is prioritized, modeling and simulation are better integrated with empirical test design, and test realism is embedded earlier and more dynamically in the acquisition lifecycle*. This would be a true structural reimagining of

how knowledge is generated, validated, and acted upon in the development of military capabilities. Under the legacy model, testing was treated as a sequential gate, positioned at the terminus of the acquisition lifecycle, where systems underwent a formal pass/fail evaluation prior to full-rate production. That model, while defensible in an era of stable threats and long program timelines, has increasingly proven misaligned with the operational and technological demands of contemporary competition, particularly in the field of advanced munitions and energetic materials, even as many in the defense R&D and test community remain institutionally wedded to it, incentivized to pursue funding continuity over rapid transition.

In contrast, a reformed test and evaluation enterprise would begin much earlier in the acquisition process – ideally at or before the point when a program receives approval to begin formal technology development (known as Milestone A) – ensuring that test considerations are integrated into the foundational stages of system architecture and concept development. Rather than deferring validation until hardware is mature and system design frozen, the new approach would treat test design as co-equal with requirements generation and systems engineering. Yet as a former director of Army test and evaluation has cautioned, while early-phase testing is a worthy goal, many programs fail at this stage because the systems in question are immature or still prototypical. Premature assumptions of test readiness can result in inaccurate data or misallocated resources, necessitating rework downstream. A reformed enterprise must balance ambition with developmental realism.[12] This shift in emphasis would place a premium on early-phase, platform-independent testing infrastructure – facilities and tools capable of evaluating materials, subsystems, or novel effects well before they are tied to a specific program of record. In the domain of energetics, for example, this would include flexible blast chambers, high-speed diagnostics suites, and instrumented range segments dedicated to characterizing new materials at various scales, from subcomponent to warhead-level. Such infrastructure could be jointly managed and supported across Services to serve as a national proving ground for lethality science.

Concurrently, modeling and simulation (M&S) would be deeply integrated into the empirical foundation of test design. Rather than using M&S as a *post hoc* justification mechanism or standalone analytical product, test planners would use validated models to generate hypotheses, identify edge cases, and determine the most efficient and informative test points.

---

[12] OSD's developmental test office, in partnership with Navy testers, proposed as recently as 2023 a framework for distributed, collaborative testing that could fill the void left by a pared-down DOT&E. It recommends treating T&E not as a one-time hurdle but as a continuum across a program's lifecycle, with integrated planning and data-sharing between developers and operators. Notably, it emphasizes maintaining independent evaluation even in a more integrated model, hopefully ensuring that while developmental and operational test events are coordinated, the assessment of results remains unbiased. The vision includes key enablers like robust Live-Virtual-Constructive (LVC) test environments and common data architectures to support joint analysis: Collins, Christopher, and Kenneth Senechal. "Test and Evaluation as a Continuum," *The Journal of Test & Evaluation* 44, no. 1 (2023).

These simulations would be grounded in real data from earlier test events – "calibrated and credibly anchored," in the language of recent OSD guidance – and would function iteratively with physical testing. The critical role of validated M&S in accelerating development and reducing empirical burden has never been clearer. Industry has already begun using digital test surrogates – such as Seek Eagle separation simulations – for early-phase design and clearance work, enabling them to reserve live-fly test points for edge cases and failure mode exploration. In some cases, extensive mission planning and PVI (pilot-vehicle interface) testing is now conducted entirely in synthetic environments, where updates to threat libraries or autonomy behaviors can be iterated rapidly. This shift reflects not a diminished commitment to rigor, but a transformation in method: tests are conducted continuously and digitally, governed by trusted models and curated databases. What is required now is a robust and universally accredited framework for validation, verification, and accreditation (VV&A), one that accommodates the tempo of modern programs and remains consistent across Services. DOT&E once served as the arbiter of this rigor; its absence necessitates a new authority, lest high-fidelity modeling become unmoored from real-world behavior and its predictive power degraded. In effect, T&E would become a data fusion enterprise, where the boundaries between synthetic and physical environments blur in service of decision-quality information.

Crucially, this reimagined system would prioritize test realism not as a final-stage filter, but as a constant and evolving characteristic of developmental and operational testing alike. Realism in this context refers not only to environmental fidelity (terrain, weather, countermeasures) but to the behavioral realism of adversary systems, including their tactics, timing, and decision processes. This would demand closer integration with the intelligence community, access to adversary threat models, and the use of AI-enabled adversary emulators in digital and live environments. The goal is not simply to verify whether a system performs as intended, but whether it contributes meaningfully to mission outcomes under plausible, contested conditions. A particularly promising frontier for this type of realism is the evaluation of collaborative autonomous weapons. Systems such as StormBreaker and emerging loyal wingman platforms, for example, rely on machine-to-machine communication, onboard autonomy, and real-time threat adaptation. Testing these systems is not a matter of validating hardware in isolation, but of evaluating behavior across swarms, environmental states, and mission contexts. This requires synthetic adversaries, contested decision trees, and high-frequency interactions that cannot be captured through legacy test constructs. To ensure these systems behave as intended safely, effectively, and reliably, the test enterprise must treat autonomy not as an add-on but as a central design axis, embedding adversary-informed scenario complexity, human-machine teaming dynamics, and fail-safe logic validation from the outset.

An enterprise with these features would shift the center of gravity of T&E from oversight to insight. It would encourage shared risk-taking, earlier failure and learning, and more dynamic

interaction among operators, developers, and evaluators. In such a system, test events become opportunities to shape systems, not merely judge them; test data becomes an asset to accelerate acquisition, not a bureaucratic hurdle to overcome. And most importantly, lethality is no longer assumed but demonstrated and improved in time to make a difference.

For the munitions and energetics community, the implications are profound. A system that can evaluate next-generation explosives, propellants, and warhead configurations before they are coupled to a specific missile or platform opens new lanes for innovation. It reduces the cost and risk of transitioning novel materials by providing validated data at the component level, and by decoupling material performance from the timelines of any single program of record. It permits the development of a modular, interchangeable library of lethal effects, each with a known performance envelope and validated safety profile, which can be selected and tailored for integration as operational needs evolve.

A reimagining of the test and evaluation (T&E) enterprise has important implications for how the enterprise must be funded. The legacy model, centered around the Office of the Director of Operational Test and Evaluation (DOT&E) as a centralized oversight body, was supported by a distinct program element in the Defense-Wide RDT&E account. That budget, which in recent years approached $377 million annually, sustained a broad suite of functions: oversight and evaluation of major defense acquisition programs, methodological development and data infrastructure, administration of joint T&E initiatives such as the Joint Test & Evaluation Program and the Joint Simulation Environment, and contracted support for independent analysis.[13] These centralized resources underwrote DOT&E's capacity to provide test integration, comparability across programs, and methodological rigor, especially for emerging technologies and novel operational concepts.

With the new directive effectively dissolving most of DOT&E's discretionary responsibilities, responsibility for many of those functions now shifts to the Services and individual program executive offices (PEOs).[14] In a more agile and distributed T&E enterprise, the Services will need enhanced resourcing for their operational test agencies (OTAs). These organizations, already operating with lean staff and aging infrastructure, will now be responsible not only for planning and executing tests but for ensuring their methodological soundness and operational relevance. Unless accompanied by real growth in OTA budgets, staff, and infrastructure, the redistribution of responsibilities risks becoming a shell game, transferring mission-critical functions without the resources to execute them. As a former Army T&E director has noted, treating test oversight

---

[13] DOT&E receives funding through Title IV of the defense appropriations act, categorized under Budget Activity 6.6: RDT&E Management Support: CRS IF10553 (2024), Table 1.

[14] While the office is one of more than a dozen organizations funded under Defense-Wide RDT&E, including DARPA and MDA, its statutory mandate for independent operational oversight distinguishes it from peer agencies: Congressional Research Service, Report IF10553 (2024), "RDT&E by Organization.

as a zero-sum exercise is untenable. Increased direct funding will be necessary to sustain instrumentation upgrades, data analysis tools, and expanded analytical staff. Just as important, targeted program-level RDT&E funding, particularly within Budget Activity (BA) 6.4 (system demonstration and validation), will need to be available to support test integration, modeling and simulation validation, and the qualification of novel subsystems, especially those incorporating advanced energetics. Without such funding, program managers will be structurally disincentivized from pursuing performance innovations that carry higher initial test burdens.

A second major area of investment will need to be common-use infrastructure. The logic of platform-independent testing depends on the availability of cross-Service testbeds capable of evaluating munitions effects, material performance, and system-level lethality in ways that are not tied to any single program of record. This implies a new round of capital investments, possibly led by the Test Resource Management Center (TRMC) or a reconstituted integration office, into facilities for high-fidelity blast diagnostics, advanced instrumentation suites, and digital test ranges. Shared modeling and simulation environments, such as the Joint Simulation Environment (JSE), will need to be sustained and expanded to support mission-level T&E under realistic conditions, including peer-adversary threat emulation and degraded command-and-control environments.

Perhaps nowhere is the need for funding mechanisms more acute than in the transition of advanced energetic materials. As highlighted in both the ETC report and the OUSD(R&E) NEP, the absence of dedicated transition funding remains a principal barrier to the adoption of higher-performing propellants and explosives. A reformed T&E ecosystem must be paired with BA 6.4 "transition readiness" resources, allocated not to individual programs but to joint S&T sponsors or the Services under a programming authority. Such funding would cover the costs of type qualification, modeling and validation, and safety certification of new materials before they are considered for programmatic integration. Ideally, BA 6.7 investments would also be available to support preplanned product improvements (P3I) that introduce next-generation energetics into existing systems when battlefield needs or production opportunities arise.

Inevitably, underlying these changes will have to be a shift in funding philosophy: from a centralized, bureaucratically protected oversight model to a distributed, capability-driven, and performance-integrated architecture prioritizing tools, infrastructure, and people, not paperwork. Resources will have to be targeted to developing the analytical capabilities required for early-phase test design, digital twin validation, and data-rich modeling environments. A renewed commitment to building the T&E workforce – especially in fields such as computational physics, weapon effects modeling, and data science – will have to replace the institutional knowledge once concentrated in DOT&E.

Of course, in the context of mounting fiscal pressure on Capitol Hill and a Department of Defense leadership committed to force modernization without topline growth, it is increasingly likely that reforms to the test and evaluation (T&E) enterprise will unfold in a constrained budgetary environment. While the vision of a reformed T&E system remains both compelling and necessary, the full realization of that vision – especially the expansion of shared test infrastructure, comprehensive modeling integration, and early-phase platform-independent validation – must be adapted to the realities of leaner resourcing. In such an environment, the result will be a triage logic: investing selectively in activities with the greatest leverage on acquisition speed, operational effectiveness, and warfighter confidence, while accepting deferments or tradeoffs in others.

Rather than launching major new capital programs or comprehensive test facility modernization, a budget-constrained Department is likely to pursue incremental, modular investments. Portable diagnostic systems, reconfigurable instrumentation packages, and deployable range augmentation tools can provide some measure of the needed capability at a fraction of the cost of large-scale infrastructure renewal, although they cannot fully replace it. For munitions and energetic materials testing, these modular assets, deployed to key facilities like China Lake or Redstone Arsenal, could enable earlier characterization and effects testing without requiring full program alignment or the burdens of system-level integration. But their effectiveness will depend not only on physical infrastructure but also on institutional safeguards to ensure that those conducting the tests are structurally independent from the competitive R&D funding enterprise, a separation essential to avoiding conflicts of interest and ensuring credibility.

Developers' critiques of legacy production-level test protocols adds further urgency to this approach. As programs seek to scale munitions production under tight budget and timeline constraints, outdated qualification regimes, many designed for Cold War-era batch manufacturing, are introducing avoidable bottlenecks. In some cases, redundant environmental or mechanical testing is being conducted on identical rounds already validated at the subsystem level, delaying delivery without improving confidence. This supports the case for revisiting qualification logic, particularly for munitions built on modular, digital design frameworks where material pedigree, structural loads, and operating profiles are already traceable across design iterations. The solution lies not in discarding rigor but in applying it where it matters most: focusing physical test events on unmodeled risks, novel behaviors, or warfighter-critical effects, and trusting validated models elsewhere.

Similarly, relatively modest investments in shared digital tools and validated physics-based models can yield significant returns. Indeed, experience shows that absent coordination, well-meaning organizations can inadvertently duplicate efforts. In one case, five groups brought high-

end radiometric gear to a single test event, straining range access and adding cost without proportionate value. Such episodes illustrate the need for better planning, common-use infrastructure, and pre-coordinated resource allocation. Instead of building new range complexes, the Department can prioritize the creation of a core digital toolkit: a library of high-fidelity simulation models, threat behavior emulators, and lethality estimation algorithms usable across Services. By embedding these tools within a federated but coordinated digital environment – managed perhaps by a reconfigured integration body – DoD can improve test design and reduce empirical burden, especially for systems integrating novel energetic effects.[15]

One of the most pragmatic adaptations in a lean environment would be the creation of small, embedded test integration cells within selected Program Executive Offices (PEOs). Staffed with experienced T&E professionals and digital engineering experts, these teams would work directly with program managers to design integrated test strategies, validate modeling assumptions, and identify empirical test points early in development. These embedded cells could be complemented by a parallel effort to centralize physical test execution at a limited number of independent, test-focused facilities, such as White Sands Missile Range, that maintain structural separation from development programs and specialize in objective, high-fidelity evaluation. While far less resource-intensive than centralized oversight structures, such teams would have a disproportionately positive impact by institutionalizing testability as a design attribute and shortening discovery timelines.

In a constrained resource environment, digital-first testing offers a path to efficiency without sacrificing rigor. Programs would submit digital test plans at Milestone reviews, linking their acquisition strategy to a validated model base and a defined empirical test campaign. The goal would not be to eliminate live testing, but to reduce its cost, duration, and sequencing friction by making it targeted and informed. Especially for advanced munitions, where full-system live tests are inherently expensive, a rigorous and accredited digital framework can preserve scientific integrity while improving speed to fielding.

Reform under constraint is not without risk. The most immediate danger is fragmentation. Without a strong center of gravity or shared integration standards, T&E practices may diverge sharply across Services, undercutting comparability, interoperability, and confidence in cross-domain effectiveness. Energetics programs are particularly at risk when funding falls short for qualification and scale-up, as this can lead to insufficient understanding of how the materials

---

[15] Former USD(A&S) William LaPlante has cautioned that even if oversight shifts to the Services, "you need that independent view" somewhere in the process and affirmed that Service test organizations alone might cut corners under schedule pressure. DOT&E's principal role, he stresses, was to ensure test plans weren't watered down. If OSD-level testing authority is diminished, alternative safeguards (e.g. strengthened Service OTAs, joint test boards, or third-party reviewers) must be put in place to preserve an independent check on test adequacy and reconstitute an "honest broker" function in whatever T&E regime emerges next.

behave across different conditions, uses, and production methods. In the absence of program-agnostic test beds or centrally sponsored transition funding, promising materials may never make it past the laboratory – not because they fail to perform, but because no entity is structurally equipped to carry them forward. There is also a serious risk that digital tools will be over-relied upon without adequate empirical grounding. Without verified, validated, and accredited models, continuously refined by high-quality test data, modeling and simulation can produce a false sense of certainty. In fields like blast physics and multi-phase detonation, where outcomes are highly sensitive to material configuration, geometry, and context, the stakes of unverified modeling are operational, not academic.

Perhaps most critically, constrained budgets will almost certainly accelerate the erosion of the independent assessment capacity that the T&E community has historically relied on. As experienced analysts and test engineers retire or are reassigned, the remaining workforce may not have the bandwidth to maintain methodological rigor or conduct adversary-informed operational realism. Without deliberate attention to highly specialized workforce development, the long-term health of the T&E enterprise may be compromised.[16]

Notwithstanding these concerns, a reformed T&E enterprise remains achievable – even under fiscal restraint – if the Department is prepared to act with focus, flexibility, and strategic clarity. Indeed, not just achievable, but imperative, lest innovation grind to a halt or the Services suddenly grow comfortable with unreliable, dangerous, or ineffective systems. Encouragingly, at least one new institution has been expressly designed to confront many of these challenges, albeit within a domain-specific scope. The recently established Joint Energetics Transition Office (JETO), created by Section 241 of the 2024 National Defense Authorization Act and embedded within the Office of the Under Secretary of Defense for Acquisition and Sustainment, reflects a Congressional recognition that the U.S. energetic materials enterprise requires systematic modernization. JETO's responsibilities span the full life cycle of energetic materials, from raw ingredient supply chain resilience to system integration and accelerated qualification. Critically, the statute tasks JETO with coordinating test and evaluation across the department, underwriting prototyping and demonstration-scale production, and establishing expedited paths for material qualification and validated modeling and simulation. Its mission – to accelerate next-generation energetics into systems – cannot succeed under the old T&E paradigm. It demands a flexible, mission-engineering-informed approach that supports iterative testing, early failure discovery, and rapid learning. Well-led and resourced, the office can be a cornerstone in the reconstruction of an evidence-driven transition pipeline, anchoring a more modular, adaptive, and test-informed approach to materials innovation and weapon system

---

[16] Former DOT&E chief scientist Dr. Greg Zacharias has pointed out that the 250+ expert testers being cut are "the cream of the crop" and that "[w]e're looking at an 80 percent reduction…They will unquestionably fail at their mission:" see Tirpak, "Cut to DoD's Test Enterprise."

lethality. As such, it may offer a partial institutional template for broader reforms, particularly those concerned with reasserting the role of T&E as a check against acquisition drift and programmatic self-dealing.

Reform, however necessary, must be guided by realism. The historical tendency to blame excessive 'oversight' requirements for program failure is often just an expedient: *the reality is that programs often fail because systems do not work as intended. The test community must remain empowered to reveal that fact, not blamed for it*. By investing in those elements of T&E that offer the greatest leverage – digital integration, early-phase realism, selective infrastructure, and embedded expertise – DoD can build on the Secretary's directive to reform what has long been a troubled T&E process while maintaining the credibility, relevance, and speed of its test architecture. More than that, it can begin to align test and evaluation not just with the pace of acquisition, but with the demands of a modern strategic environment in which lethality, reliability, and rapid delivery are essential and inseparable.

Recommendations:
Aligning the Enterprise with the New Paradigm

The following recommendations are organized into four interrelated categories that collectively define the foundations of a reformed and future-ready T&E enterprise. The categories reflect not only the key areas of vulnerability in the wake of the May 2025 directive but also the principal levers by which Congress, the Services, and OSD can reconstitute a T&E function that is empirically grounded, technologically modern, and operationally relevant. Taken together, these recommendations are intended to secure the autonomy, competence, and integrity of T&E in support of lethality, survivability, and warfighter trust.

Institutional Independence

- Preserve DOT&E's statutory independence, but realign its role around continuous evaluation, data trend analysis, and independent sufficiency review, as reflected in the FY26 NDAA's proposed Alternative T&E Pathway.
- Support the expansion of the Alternative T&E Pathway beyond software programs to include energetics-intensive systems, complex munitions, and integrated weapon effects testing.
- Require that operational test results be reported in unclassified form to Congress without alteration or filtering by the Services or acquisition authorities.
- Establish a standing congressional reporting requirement for any waivers granted under Middle Tier Acquisition or Rapid Fielding pathways that bypass operational test phases.

Digital Infrastructure

- Accelerate investment in simulation environments (e.g., Joint Simulation Environment) and associated VV&A frameworks to support next-generation system testing.
- Require M&S tools used in OT&E to undergo third-party verification and threat-informed accreditation to ensure they reflect current adversary capabilities and dynamic threat evolution.
- Fund a shared digital infrastructure for empirical data archiving and reuse across programs and Services to support longitudinal learning and threat-based test design.

Workforce and Resourcing

- Provide stable and independent appropriations for the operational test enterprise, insulated from the influence of Service acquisition programs.

- Direct OSD(P&R) and USD(A&S) to jointly develop a Test and Evaluation Workforce Plan to address gaps in cyber survivability, digital engineering, and mission engineering expertise.
- Expand the Defense Civilian Training Corps and related programs to create a pipeline of operationally literate, technically proficient testers.

Methodological Reform

- Require early and continuous test data collection throughout development, including from prototypes and live-fire events, with evaluation strategies that evolve dynamically in response to emerging results and operational priorities.
- Require live-fire and survivability testing for all new munitions and platforms involving novel materials, warheads, or propulsion technologies, regardless of acquisition pathway.
- Promote integrated T&E frameworks that link platform testing to mission- and campaign-level effectiveness assessments using empirical data.
- Restore and strengthen the Joint Technical Coordinating Group for Munitions Effectiveness (JTCG/ME) and Joint Test & Evaluation (JT&E) programs as centers of methodological innovation.
- Encourage the development of unified test and evaluation strategies for all covered programs, aligning developmental and operational objectives, and minimizing duplicative or milestone-specific test events.