WEAPON SYSTEMS TESTING

DOD Needs to Update Policies to Better Support Modernization Efforts

Report to Congressional Committees

December 2025

GAO-26-107009

United States Government Accountability Office

Accessible Version

GAO Highlights

WEAPON SYSTEMS TESTING

DOD Needs to Update Policies to Better Support Modernization Efforts

GAO-26-107009

December 2025

A report to congressional committees.

For more information, contact: Shelby S. Oakley at oakleys@gao.gov.

What GAO Found

The Department of Defense (DOD) identified test and evaluation modernization as a crucial part of its effort to get capabilities to warfighters faster. DOD organizations, including the Office of the Secretary of Defense and the military departments, have undertaken modernization planning efforts with varying areas of focus and levels of detail. Nonetheless, these plans share themes, including the use of digital engineering tools and highly skilled workforces.

GAO's analysis of **DOD-wide test and evaluation policies** found they were not fully consistent with selected leading practices for product development as applied to test and evaluation: involve testers early, conduct iterative testing, use digital twins and threads, and obtain user feedback iteratively. These policies contained some tenets of the leading practices, particularly for the software acquisition and urgent capability acquisition pathways. However, these leading practices were largely not reflected in the policies for programs in the major capability acquisition and middle tier of acquisition pathways, which account for the majority of DOD spending on weapon systems acquisition.

Key DOD-wide Weapon Systems Testing Policies Fall Short of Selected Leading Practices

Leading practices for product development applied to test and evaluation	Department of Defense (DOD) test and evaluation policy			
	Current developmental test policy		Current operational and live-fire test policy	
	Overall	MCA and MTA	Overall	MCA and MTA
Involve test officials in development of acquisition strategy	◐	MCA ○ MTA ◐	◐	◐
Conduct iterative testing to generate minimum viable product	◐	○	◐	○
Use digital twins and digital threads	◐	○	◐	◐
Obtain user feedback throughout iterative testing	◐	○	◐	○

◐ Policy partially reflects leading practice ○ Policy does not reflect leading practice

MCA = Major capability acquisition
MTA = Middle tier of acquisition
} Majority of DOD spending on weapon systems acquisition

Overall = overall assessment of test policy for four weapon system acquisition pathways: MCA, MTA, urgent capability acquisition, and software acquisition

Source: GAO analysis of DOD-wide policies applicable to test and evaluation. | GAO-26-107009

Accessible Data for Key DOD-wide Weapon Systems Testing Policies Fall Short of Selected Leading Practices

Department of Defense (DOD) test and evaluation policies

Current developmental test policy		Current operational and live-fire test policy	
Overall	**MCA and MTA**	**Overall**	**MCA and MTA**
Partially reflects	MCA does not reflect; MTA partially reflects	Partially reflects	Partially reflects
Partially reflects	Does not reflect	Partially reflects	Does not reflect
Partially reflects	Does not reflect	Partially reflects	Partially reflects
Partially reflects	Does not reflect	Partially reflects	Does not reflect

Source: GAO analysis of DOD-wide policies applicable to test and evaluation. | GAO-26-107009

Further, GAO found that **DOD's digital engineering policy** and the test and evaluation section of **DOD's systems engineering policy** do not describe specific processes to ensure application of leading practices to testing.

GAO also found that **military department-level test and evaluation policies** generally did not reflect the leading practices beyond the level found in DOD-wide policies. GAO similarly found that these leading practices were not reflected in key program documents, like acquisition strategies and test strategies, for selected weapon systems acquisition programs it reviewed.

DOD has a unique opportunity to not only retool its existing test and evaluation enterprise, but to redefine the role that enterprise can play in enabling faster delivery of relevant capabilities to warfighters. Fully incorporating leading practices into policies relevant to weapon system test and evaluation could help pivot the test enterprise's current reactive role to a proactive one, informing and aiding defense acquisition efforts.

Why GAO Did This Study

DOD has yet to realize its goal to rapidly develop weapon systems to get capabilities to the warfighter when needed. DOD acquisition programs have identified challenges discovered during test and evaluation as contributing to delays in development.

A House committee report includes a provision for GAO to assess how DOD is modernizing weapon system test and evaluation. GAO's report (1) describes DOD's plans to modernize test and evaluation to deliver capabilities faster to the warfighter, and (2) assesses the extent to which DOD-wide and military department policies for test and evaluation reflect selected GAO leading practices for product development.

To do this work, GAO assessed DOD test and evaluation modernization plans and policies and weapon system acquisition documentation. GAO visited three military department test organizations to observe tools in practice. GAO also interviewed DOD and military department officials from test organizations and other entities.

What GAO Recommends

GAO is making 13 recommendations, including that the Secretary of Defense and the Secretaries of the Air Force, Army, and Navy each should revise weapon system test and evaluation policies and other related policies to reflect selected leading practices for product development. Specifically, revisions should require involvement of testers in acquisition strategies; iterative approaches to testing, including use of digital twins and threads; and ongoing end user input. DOD concurred with seven recommendations, partially concurred with five recommendations, and did not concur with one recommendation. GAO continues to believe all 13 of its recommendations are valid, as discussed in this report.

Contents

GAO Highlights ii

What GAO Found ii

Why GAO Did This Study iii

What GAO Recommends iii

Letter 1

Background 2

Plans for Modernizing Test and Evaluation Vary Across DOD 7

Test and Evaluation Modernization Is Constrained by Policies That Do Not Fully Reflect Leading Practices 13

Conclusions 30

Recommendations for Executive Action 30

Agency Comments and Our Evaluation 32

Appendix I: Objectives, Scope, and Methodology 38

Appendix II: Comments from the Department of Defense 41

Accessible Text for Appendix II: Comments from the Department of Defense 52

Appendix III: GAO Contact and Staff Acknowledgments 61

Tables

Table 1: Relevant Leading Practices for Product Development Applied to Department of Defense (DOD) Test and Evaluation 7

Table 2: Selected DOD-wide and Military Department Plans for Test and Evaluation Modernization 8

Figures

Key DOD-wide Weapon Systems Testing Policies Fall Short of Selected Leading Practices ii

Accessible Data for Key DOD-wide Weapon Systems Testing Policies Fall Short of Selected Leading Practices iii

Figure 1: Adaptive Acquisition Framework Pathways Relevant to Weapon Systems Test and Evaluation 3

Figure 2: Test and Evaluation Organizations Across the Department of Defense (DOD) 4

Figure 3: Leading Companies Progress Through Iterative Design, Validation, and Production Cycles to Develop a Minimum Viable Product 6

Figure 4: Key DOD-wide Policies Applicable to Weapon System Test and Evaluation Do Not Fully Reflect Selected Leading Practices 15

Accessible Data for Figure 4: Key DOD-wide Policies Applicable to Weapon System Test and Evaluation Do Not Fully Reflect Selected Leading Practices 16

Figure 5: Military Department Weapon System Test and Evaluation Policies Do Not Fully Reflect Selected Leading Practices Beyond the Level in DOD-Wide Policy 19

Figure 6: T-7A Acquisition Strategy Does Not Describe Timely Access to Contractor Test Data and Digital Model 21

Accessible Data for Figure 6: T-7A Acquisition Strategy Does Not Describe Timely Access to Contractor Test Data and Digital Model 21

Figure 7: Test Plans for Air Force's F-22 SeE Program Do Not Consider How Ongoing User Feedback Could Be Incorporated into an Iterative Test Process 23

Accessible Data for Figure 7: Test Plans for Air Force's F-22 SeE Program Do Not Consider How Ongoing User Feedback Could Be Incorporated into an Iterative Test Process 24

Figure 8: Army's FLRAA Program Does Not Have an Iterative Test Approach 26

Accessible Data for Figure 8: Army's FLRAA Program Does Not Have an Iterative Test Approach 26

Figure 9: The Navy's Ford Class Aircraft Carrier Program Did Not Fully Implement an Iterative Test Process 28

Accessible Data for Figure 9: The Navy's Ford Class Aircraft Carrier Program Did Not Fully Implement an Iterative Test Process 28

Abbreviations

- AFOTEC Air Force Operational Test Center
- AFTC Air Force Test Center
- ATEC Army Test and Evaluation Command
- DOD Department of Defense
- DODI Department of Defense Instruction
- DOT&E Office of the Director, Operational Test and Evaluation
- DSB Defense Science Board
- F-22 SeE F-22 Sensor Enhancements
- FLRAA Future Long Range Assault Aircraft
- MCA major capability acquisition
- MTA middle tier of acquisition
- OPTEVFOR Operational Test and Evaluation Force
- OSD Office of the Secretary of Defense
- OUSD(A&S) Office of the Under Secretary of Defense for Acquisition and Sustainment
- OUSD(R&E) Office of the Under Secretary of Defense for Research and Engineering

This is a work of the U.S. government and is not subject to copyright protection in the United States. The published product may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.

December 11, 2025

Congressional Committees

For more than 20 years, we have reported annually on the Department of Defense's (DOD) most expensive weapon system acquisition programs, which often cost significantly more than planned, fall years behind schedule, and deliver less capability in the end than initially promised. At the same time, changes in the global threat environment and leaps in technology have changed the nature of the capabilities that DOD seeks to acquire. Weapon systems are increasingly cyber-physical—complex networks of hardware and software—with software driving programs more than ever before. Our recent work on leading practices for product development identified that delivering these complex systems with speed requires new, iterative approaches for development, including test and evaluation.[1]

Still, many weapon system acquisition programs continue to use a slow, linear approach to development and fall short of delivering relevant capabilities quickly and at scale. In June 2025, we found that DOD's expected time frame to deliver initial capabilities for 79 of its major defense acquisition programs averaged about 12 years.[2] This average was longer than the average time frame of 10 years we identified for 82 programs in May 2019—less than 1 year before DOD revamped its acquisition policy to emphasize speed.[3] In part due to these longstanding issues, DOD weapon systems acquisition has been on our High-Risk List for 35 years.[4]

Weapon system acquisition programs frequently cite challenges discovered during test and evaluation as a cause for why they cannot deliver capabilities faster. As our prior work has found though, test and evaluation of new capabilities is often just the "canary in the coal mine" providing indication on whether a program is on sound footing. Specifically, weapon system acquisition programs are regularly framed by unrealistic business cases, including optimistic assumptions by the acquisition community as to how the program will perform in test and evaluation. Later, as development progresses, and test and evaluation reveals previously-unknown deficiencies, these unrealistic business cases deteriorate, resulting in cost growth, schedule delays, and capability reductions. While this traditional, reactive test and evaluation helps warfighters understand any limitations on effectiveness and suitability associated with their weapons, it does little to aid faster development and delivery of relevant capabilities that those warfighters need.

A House committee report accompanying a bill for the James M. Inhofe National Defense Authorization Act for Fiscal Year 2023 contains a provision for us to assess how DOD is modernizing weapon systems test and evaluation to improve performance of defense acquisition programs. Our report (1) describes plans DOD has

[1]GAO, *Leading Practices: Iterative Cycles Enable Rapid Delivery of Complex, Innovative Products*, GAO-23-106222 (Washington, D.C.: July 27, 2023).

[2]GAO, *Weapon Systems Annual Assessment: DOD Leaders Should Ensure that Newer Programs Are Structured for Speed and Innovation*, GAO-25-107569 (Washington, D.C.: June 11, 2025).

[3]GAO, *Weapon Systems Annual Assessment: Limited Use of Knowledge-Based Practices Continues to Undercut DOD's Investments*, GAO-19-336SP (Washington, D.C.: May 7, 2019).

[4]GAO, *High-Risk Series: Heightened Attention Could Save Billions More and Improve Government Efficiency and Effectiveness*, GAO-25-107743 (Washington, D.C.: Feb. 25, 2025).

established to modernize test and evaluation to deliver relevant capabilities faster to warfighters, and (2) assesses the extent to which DOD-wide and military department policies for test and evaluation reflect our selected leading practices for product development.

To address our objectives, we reviewed and assessed DOD documentation, including test and evaluation policies, other related functional area policies, test and evaluation modernization plans and strategies, and weapon system acquisition documentation. We assessed the DOD-wide and military department test and evaluation policies and other related policies against our leading practices for product development. We focused our review on test and evaluation policies, rather than guidance, because these policies require courses of action while the test and evaluation guidance is generally not binding and therefore optional to follow. We conducted three site visits to the Air Force Test Center (AFTC), Army Test and Evaluation Command (ATEC), and Navy Operational Test and Evaluation Force (OPTEVFOR). We selected these sites because they are primary test agencies at each of the military departments and have centrally located workforces to meet with test officials and see test and evaluation tools in practice, such as modeling and simulation and live virtual constructive environments.

Additionally, we interviewed officials from DOD-wide and military department-level test organizations and other entities to obtain additional context and corroborate our document reviews and assessments. Those organizations and entities include the Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)); Office of the Director, Operational Test and Evaluation (DOT&E); the Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD(A&S)); Air Force Operational Test Center (AFOTEC); AFTC; ATEC; the Office of the Chief of Naval Operations; and OPTEVFOR. We also reviewed program documentation for eight selected ongoing weapon system acquisition programs to examine the military departments' implementation of selected leading practices for product development. Appendix I presents a detailed description of the objectives, scope, and methodology for our review.

We conducted this performance audit from August 2023 to December 2025 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

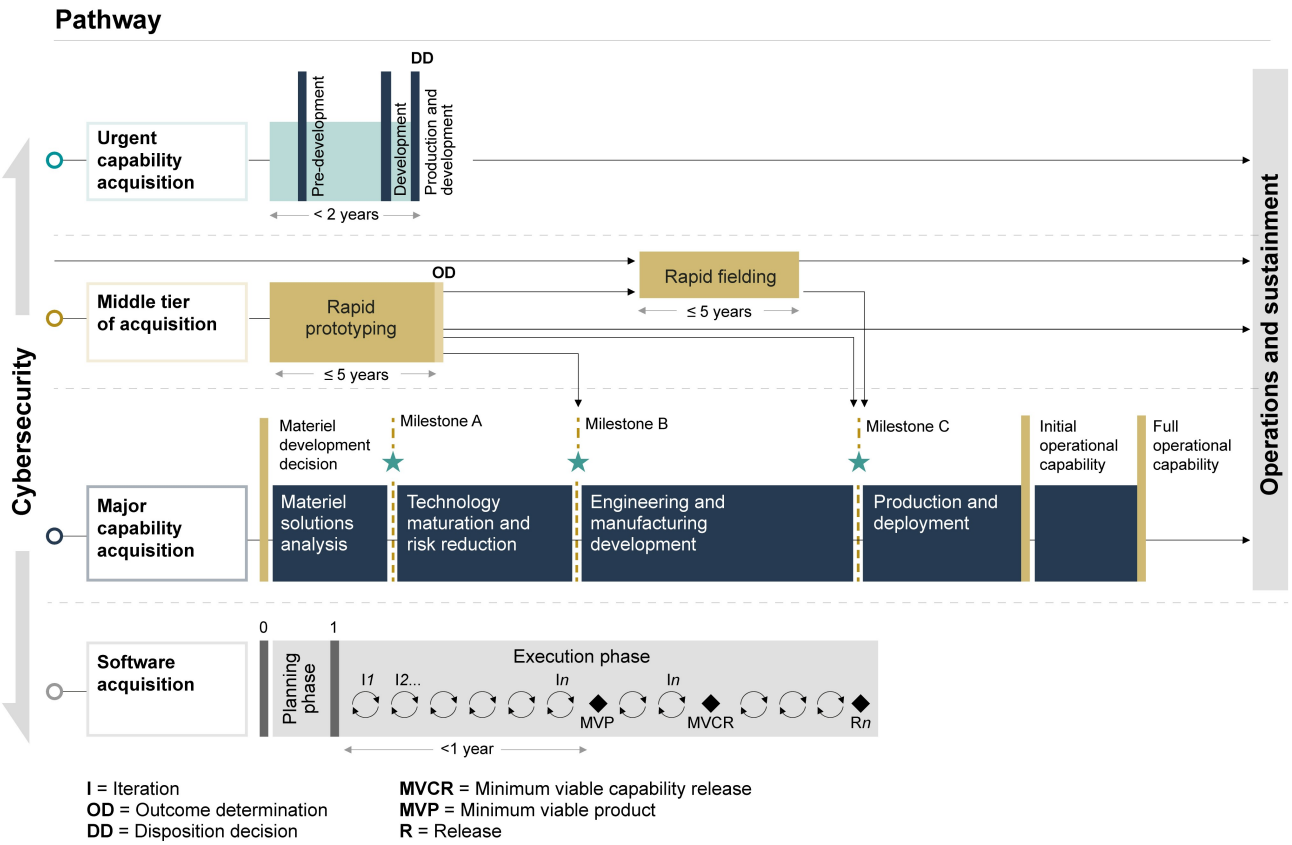
Background

DOD's Adaptive Acquisition Framework

In January 2020, DOD revised its acquisition policies as part of an effort to deliver more effective, suitable, survivable, sustainable, and affordable solutions to the warfighter in a timely manner. These changes were designed to allow military departments to select one or more pathways for acquiring a weapon system to best match the capability being acquired, as well as tailor and streamline certain processes. The Adaptive Acquisition Framework emphasizes several principles that include simplifying acquisition policy, tailoring acquisition approaches, and conducting data-driven analysis. The framework includes six acquisition pathways, but for the purposes of this report, we reference only the four acquisition pathways relevant to test and evaluation of weapon systems: (1) urgent capability acquisition, (2) middle tier of acquisition, (3) major capability acquisition, and (4) software acquisition. Each pathway has unique processes, reviews,

documentation requirements, and metrics designed to match the characteristics and risk profile of the capability being acquired (see fig. 1).

Figure 1: Adaptive Acquisition Framework Pathways Relevant to Weapon Systems Test and Evaluation



Source: GAO analysis of Department of Defense (DOD) data. I GAO-26-107009

Overview of Weapon Systems Test and Evaluation Activities and Oversight

Test and evaluation is a critical aspect of ensuring successful outcomes for weapon systems throughout their life cycles. Testing provides opportunities to collect data on system performance and identify and resolve system errors before a final fielding or acquisition decision is made. DOD conducts various types of testing activities:

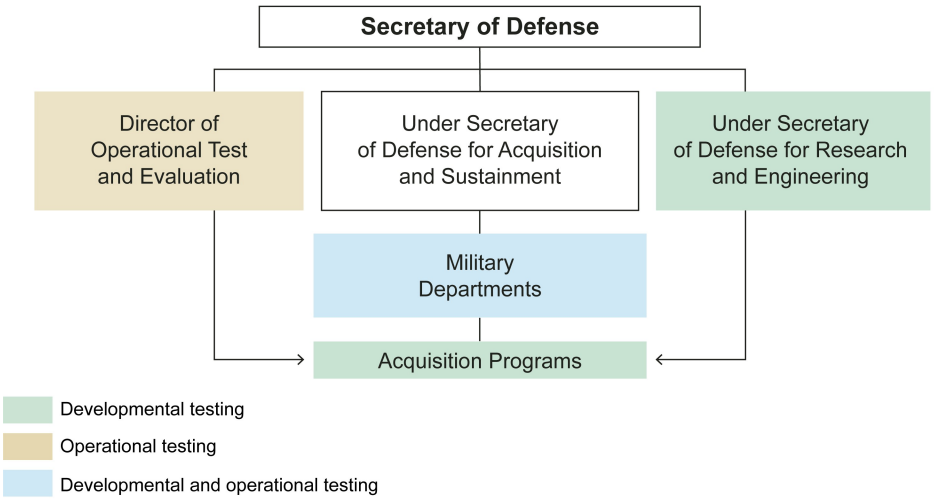
- **Developmental testing:** conducted by contractors and the government and designed to provide feedback on a system's design and combat capabilities before initial production or deployment.
- **Operational testing:** conducted by the government and designed to evaluate a system's effectiveness and suitability to operate in realistic conditions before full-rate production or deployment. Traditionally, DOD has completed initial operational test and evaluation using an acquisition program's low-rate initial production quantities.

- **Live-fire testing:** conducted by the government and involves the realistic survivability and lethality testing of systems configured for combat.

DOD test and evaluation policy also emphasizes the importance of integrated testing.[5] This type of testing is built on collaboration between developmental and operational test officials and involves the sharing of data and planning resources to facilitate communication and allow testing to build on previous efforts. Integrated testing can help identify deficiencies in a system design and help inform corrective fixes early during development and potentially reduce the scope of dedicated operational testing at the end of the acquisition process.

The Office of the Secretary of Defense (OSD) and the military departments share oversight responsibilities for test and evaluation (see fig. 2).

Figure 2: Test and Evaluation Organizations Across the Department of Defense (DOD)



Source: GAO presentation of Department of Defense information. | GAO-26-107009

OUSD(R&E) and DOT&E have the statutory authority to oversee test adequacy and establish policies for developmental testing activities and operational testing activities, respectively.[6] Beyond policy, OSD and the military departments can further define test and evaluation processes by issuing guidance. The military departments have officials in their own test and evaluation organizations and in acquisition programs who execute testing.

- **DOT&E** is the organization responsible for oversight of operational testing and live-fire testing across the department. The office has broad statutory authority to create policies and monitor how operational testing is conducted in the enterprise, as well as monitor and review all live-fire test activities.[7] Statute also requires DOT&E to approve plans for all operational test and evaluation of major defense acquisition

[5]Integrated testing does not replace or eliminate the requirement for initial operational test and evaluation, as a condition to proceed beyond low-rate initial production as required by 10 U.S.C. § 4171. Department of Defense Instruction 5000.89, *Test and Evaluation* (Nov. 19, 2020).

[6]10 U.S.C. §§ 133a, 139.

[7]10 U.S.C. § 139.

programs.[8] DOD acquisition and test and evaluation policy further defines oversight responsibilities for approving operational testing in program test plans and strategies along with assessing the adequacy of a program's test execution for weapon system programs. In a May 2025 memorandum, the Secretary of Defense directed DOT&E to eliminate any non-statutory or redundant functions and conduct a civilian reduction-in-force.[9]

- **OUSD(R&E)** develops governing policies relating to research and engineering, technology development, prototyping, and all developmental test activities and programs.[10] Within OUSD(R&E), the Office of Developmental Test Evaluation and Assessments oversees, in collaboration with other test and evaluation stakeholders, the development of test plans, including test activities, execution, and evaluation.
- **Military departments** have their own test and evaluation organizations that develop test plans and conduct developmental and operational testing, working in coordination with the acquisition programs, contractors, and warfighters, among others. Each military department assigns responsibilities for different test activities differently across its test organizations and acquisition programs.
- **Acquisition programs** can also have their own developmental test officials who conduct testing or review test data from the contractor.

In November 2025, the Secretary of Defense issued a memorandum to transform the defense acquisition system into the warfighting acquisition system to accelerate the fielding of urgently needed capabilities.[11] As part of the transformation, the memorandum directs the department to modernize test infrastructure, reduce test oversight, implement an adaptable test approach, and streamline test and evaluation requirements. DOD issued this memorandum several months after we provided a draft of this report to DOD for comment. As a result, we did not specifically assess how the changes called for by this memorandum will address the challenges identified in our report.

Leading Practices for Product Development

In July 2023, we reported how leading companies rely on iterative cycles to deliver innovative cyber-physical products with speed.[12] This iterative development structure—a process that includes test and evaluation activities—involves continuous cycles that include common key practices, such as identifying a minimum viable product and obtaining user feedback to ensure that capabilities are relevant and responsive to users' needs. The iterative structure is enabled by digital engineering, including the use of digital twins or digital threads.[13]

Activities in these iterative cycles often overlap as the design undergoes continuous user engagement and testing. We found that as the cycles proceed, leading companies' product teams refine the design to achieve a minimum viable product—one with the initial set of capabilities needed for customers to recognize value

[8]10 U.S.C. § 4171.

[9]Secretary of Defense Memorandum, *Reorganization of the Office of the Director of Operational Test and Evaluation* (May 27, 2025).

[10]10 U.S.C. § 133a.

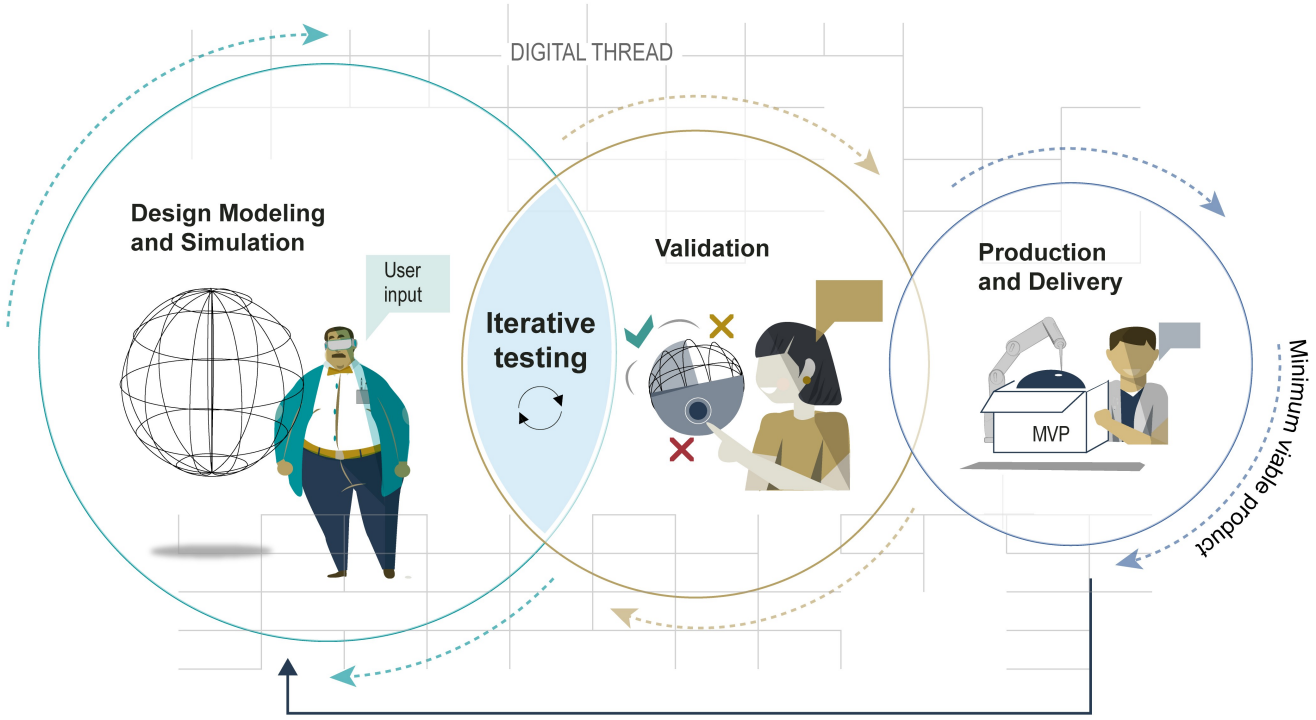
[11]Secretary of Defense Memorandum, *Transforming the Defense Acquisition System into the Warfighting Acquisition System to Accelerate Fielding of Urgently Needed Capabilities to Our Warriors* (Nov. 7, 2025).

[12]GAO-23-106222.

[13]Digital twins are virtual representations of physical products that incorporate dynamic data regarding a physical object or a system—meaning the model changes and updates in real time as new information becomes available. Digital threads are common sources of information that connect stakeholders with real-time data across the product life cycle to help inform decisions.

suitable to field and can be followed by successive iterations. These companies use modern design and manufacturing tools and processes to produce and deliver the product in time to meet their customers' needs. Figure 3 depicts key elements of this approach.

Figure 3: Leading Companies Progress Through Iterative Design, Validation, and Production Cycles to Develop a Minimum Viable Product



Source: GAO analysis of leading company information; GAO (illustration). | GAO-26-107009

Over the past few years, we have recommended that DOD and the military departments update their pathway policies to incorporate our leading practices for product development.[14] DOD is updating its Adaptive Acquisition Framework policies, but one recently updated policy and the draft we reviewed of another policy do not fully implement leading practices to achieve positive outcomes. Further, although DOD's military departments have issued policies in alignment with DOD's goals and the Adaptive Acquisition Framework, they do not consistently reflect leading practices.[15]

Table 1 details how leading practices for product development can be applied to test and evaluation to support rapid development of weapon systems.

[14]GAO, *DOD Acquisition Reform: Military Departments Should Take Steps to Facilitate Speed and Innovation*, GAO-25-107003 (Washington, D.C.: Dec. 12, 2024); *Middle-Tier Defense Acquisitions: Rapid Prototyping and Fielding Requires Changes to Oversight and Development Approaches*, GAO-23-105008 (Washington, D.C.: Feb. 7, 2023); and *Leading Practices: Agency Acquisition Policies Could Better Implement Key Product Development Principles*, GAO-22-104513 (Washington, D.C.: Mar. 10, 2022).

[15]GAO-25-107743.

Table 1: Relevant Leading Practices for Product Development Applied to Department of Defense (DOD) Test and Evaluation

Leading practices [category]	Leading practices [subcategory]	Application of leading practice to test and evaluation
Involvement in early acquisition decisions	Employ and empower right-sized teams of multi-disciplined stakeholders	Involve test officials early in defining the development program, including in the development of the acquisition strategy and by providing access to the key items, such as digital twins and digital threads, needed to conduct testing.
Iterative development	Implement iterative testing processes to generate a minimum viable product	Develop test plans that reflect an iterative testing approach and integrate developmental and operational testing to support rapid development of a minimum viable product.
Iterative development	Use modern design tools, including digital twins and digital threads	Ensure testers use key modern design tools, including digital twins and digital threads developed by the design contractor and needed to support iterative government testing.
User feedback	Actively engage users to obtain feedback throughout an iterative development process	Develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input throughout testing to ensure the capability under development is relevant and responsive to users' most critical needs.

Source: GAO analysis of leading practices on product development applied to DOD test and evaluation processes. | GAO-26-107009

Plans for Modernizing Test and Evaluation Vary Across DOD

DOD has identified test and evaluation modernization as a crucial part of its larger effort to deliver new capabilities with speed. Because test and evaluation responsibilities are divided among different organizations within DOD—including at the Office of the Secretary of Defense and military department levels—these planning efforts have varied in terms of their areas of focus and their level of detail, among other differences. Even so, multiple test organizations identified several common themes that were significant to their plans for modernizing test and evaluation, including increased use of digital twins and digital tools and highly skilled test and evaluation workforces.

DOD and Military Departments Describe Modernization Efforts for Test and Evaluation in Various Plans

DOD-wide and military department-level organizations describe their plans for modernizing test and evaluation in various ways, such as strategic plans, task force briefings, and other documents. These plans vary in their focus, the level of detail and number of action items presented. For example, DOT&E's Strategy Implementation Plan and its Science and Technology Strategic Plan outline strategic goals and objectives for modernizing test and evaluation. Table 2 describes selected DOD-wide and military department plans for modernizing test and evaluation that we reviewed.

Table 2: Selected DOD-wide and Military Department Plans for Test and Evaluation Modernization

Category	Plan	Organization	Description
DOD-wide plans	Test and Evaluation, August 2024	Defense Science Board (DSB)	A DOD-wide report tasked by the Under Secretary of Defense for Research and Engineering, in response to a congressional directive in statute to evaluate DOD's test and evaluation activities and explore opportunities to increase test and evaluation speed and efficiency. The report provides several findings and recommendations calling for a strategic shift in DOD acquisitions and identified key factors related to test and evaluation for achieving this goal.
DOD-wide plans	Digital Engineering Capability to Automate Testing and Evaluation Final Report, May 2024	DSB	Final report from the congressionally directed DSB Task Force on Digital Engineering Capability to Automate Testing and Evaluation. The task force found that proper application of digital engineering can improve cost, schedule, and performance of complex projects and programs. They recommended, among other things, that DOD-wide organizations and the military services invest in digital engineering infrastructure and workforce and accelerate the use of digital engineering in testing.
DOD-wide plans	Director, Operational Test & Evaluation (DOT&E) Strategy Implementation Plan, April 2023	DOT&E	A DOD-wide plan that outlines five strategic pillars: test the way we fight; accelerate the delivery of weapons that work; improve the survivability of DOD in a contested environment; pioneer test and evaluation of weapon systems built to change over time; and foster an agile and enduring test and evaluation enterprise workforce. These pillars are intended to produce end states such as testing under operational environment conditions, using digital tools and data to accelerate weapons delivery, and having a highly skilled test and evaluation workforce.
DOD-wide plans	DOT&E Science and Technology Strategic Plan, January 2021	DOT&E	A DOD-wide plan that documents DOT&E's framework for addressing key science and technology priorities including test and evaluation for software and cybersecurity, next-generation test and evaluation capabilities, the integrated test and evaluation life cycle, digital transformation, and workforce expertise.
DOD-wide plans	Test and Evaluation as a Continuum, March 2023	Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E))	A white paper by a test official within OUSD(R&E) that calls for a new integrated framework of test and evaluation activities using capability and outcome-focused testing, an agile and scalable evaluation framework, and enhanced test design. This framework contrasts with the current serial nature of test activities. An OUSD(R&E) official told us they are developing a communications strategy to socialize this framework with the rest of DOD's test and evaluation community and are identifying programs to serve as pathfinders of these practices. This framework is not included in policy nor adopted in practice across DOD.
Military department-level plans	Fiscal Year 2024 Army Test and Evaluation Command (ATEC) Next: Campaign Plan for 2024-2029	ATEC	A plan for the Army that sets multi-year objectives and identifies specific fiscal year 2024 tasks related to modernizing the workforce, investing in test and evaluation capabilities, and leveraging digital technologies, among other things.

Category	Plan	Organization	Description
Military department-level plans	U.S. ATEC Modernization Task Force Summary, March 2024	ATEC	Briefing slides that summarize focus areas—including workforce transformation, business processes, and test and evaluation capabilities.
Military department-level plans	Air Force Operational Test and Evaluation Center (AFOTEC) Strategic Vision 2030, February 2024	AFOTEC	Briefing slides that summarize AFOTEC's strategic vision and progress made toward identified end states. Key issue areas include talent management, policies, tools, and integrated test capabilities.
Military department-level plans	Navy Operational Test and Evaluation Force (OPTEVFOR) Strategic Roadmap, January 2022	OPTEVFOR	A roadmap to define strategic priorities and measure progress of each initiative, based on the Chief of Naval Operations' Navigation Plan and its Implementation Framework. Focus areas include implementing adaptive relevant testing, external engagement, improving cyber testing, and command operations.

Source: GAO summary of Department of Defense (DOD) and military department information. | GAO-26-107009

Note: GAO also reviewed DOD-wide and military department plans, such as strategies for implementing digital engineering, that were not directly about modernizing test and evaluation but could affect test practices.

In response to a congressional directive in statute, OUSD(R&E) directed the Defense Science Board (DSB)—a DOD committee that provides independent advice and recommendations on matters supporting scientific and technical enterprise—to report on how DOD can improve the speed and efficiency of test and evaluation. Congress also directed DSB to assess DOD's progress in implementing a digital engineering capability to automate testing and evaluation.[16] The reports describe a DOD-level end-state of modernized test and evaluation. DOD is in the process of responding to DSB's recommendations. OUSD(R&E)'s Office of Developmental Test Evaluation & Assessment plans to address relevant recommendations through updated developmental test policy, which the office is currently revising. Further, a DOT&E official stated that DOT&E does not have a formal response to the digital engineering report and that the focus areas identified in the test and evaluation report reflected issues of which the test community was already aware. According to the DOT&E official, that office is already implementing the recommendations from the DSB reports through its Strategy Implementation Plan.

Additionally, some of these plans listed in table 2 outlined conceptual frameworks. For example, an OUSD(R&E) official described their "Test and Evaluation as a Continuum" concept as a framework to move the test enterprise into a more iterative structure. The "Test and Evaluation as a Continuum" concept—which is not included in policy and is not adopted in practice—is a pivot from DOD's traditional linear testing and systems engineering approach.

Specifically, in DOD's traditional approach, DOD begins testing a weapon system after sufficient requirements and design are set. In contrast, the "Test and Evaluation as a Continuum" concept moves test and evaluation from a serial set of activities conducted largely after system design to a continuum of activities beginning once a mission need is identified. A model-based environment will enable mission engineering, systems engineering, and test and evaluation communities to work in tandem early and continually and to respond

[16]National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116-92, § 231 (2019).

quickly to the dynamic threat environment.[17] The framework also states that in the future, the "requirements" of a particular weapon system will most likely evolve from discrete key performance parameters to one of "system behaviors" where the continuous testing process will allow more rapid evaluation against those desired system behaviors within the context of the model-based environment. Further, OUSD(R&E) expects that the missions engineering process associated with "Test and Evaluation as a Continuum" will identify mission sets as well as the capabilities needed to conduct them, thereby enabling timely updates to performance requirements for the weapon system under development.

Because the modernization plans differ in how much detail they provide regarding steps the organization plans to take, we did not assess the degree to which these plans have been implemented. However, officials from DOD-wide and military department-level test organizations described various efforts taken to modernize test and evaluation in accordance with relevant plans, with implementation in various stages. For example, OUSD(R&E) is currently developing tenets for "Test and Evaluation as a Continuum" while some military departments developed targets to help them measure progress toward meeting their goals.

Plans Share Several Focus Areas to Modernize Test and Evaluation

Despite the varied purpose, focus, and level of detail, we found several common themes across the modernization plans we reviewed, including but not limited to investment in test infrastructure, workforce transformation, and revisions to policy and guidance.

Investment in Test Infrastructure

Plans from each of the DOD-wide and military department test organizations, as well as the DOD Strategic Management Plan, identify the need to invest in test infrastructure.[18] Specifically, multiple plans call for infrastructure investments to increase the use of digital engineering and business tools, such as:

- **Modeling and simulation and digital twins.** Plans from each military department, DOT&E, and OUSD(R&E) call for expanded use of modeling and simulation and digital engineering tools. For example, DOT&E's Strategy Implementation Plan calls for increasing the use of credible digital twins in test and evaluation. Further, the DSB report on test and evaluation recommends that programs be prepared to execute test and evaluation using digital engineering principles and to ensure they develop and deliver digital tools to enhance testing strategies.

  At the military department-level, ATEC—the Army's developmental and operational test organization—describes in its 2024-2029 Campaign Plan its goal for digital transformation that leverages a distributed test network. Currently, ATEC digitally models its vehicle test ranges, which allows it to simulate various conditions and run digital simulations to collect more data than would be possible with live testing. ATEC officials told us that some of the advantages of these digital simulations were that they could run far more

[17]Mission engineering is an interdisciplinary approach and process encompassing the entire technical effort to analyze, design, and integrate current and emerging operational needs and capabilities to achieve desired mission outcomes. Systems engineering is a methodical and disciplined approach for the specification, design, development, realization, technical management, operations, and retirement of a system.

[18]Infrastructure to support test and evaluation can include a variety of physical and digital investments including test ranges and equipment, facilities, and various digital tools.

iterations of a test than is possible on a physical range, including simulating changes in weather conditions that may not be present on a physical range at the time of test.

- **Digital environments and distributed test networks.** DOT&E's Strategy Implementation Plan calls for developing the requirements and concept for a digital environment to integrate with live testing.

  At the military department-level, AFOTEC—the Air Force's operational test agency—calls for a digital transformation of processes and infrastructure to enable operational testing and business activities in a digital environment. Additionally, ATEC's Modernization Task Force identified a need to connect the ATEC enterprise on a distributed test network and to perform data collection, instrumentation, analysis, and reporting in a cloud-based environment. Leveraging this network, ATEC officials conducted a test event using a mixture of live and virtual assets across the country.

- **Digitized business tools**. Multiple test organizations also identified improvements to business tools as an area of modernization. For example, DOT&E's Science and Technology Plan and Navy officials we interviewed highlighted the use of a digital test plan to streamline their business processes by reducing the development time of the test plan and fully integrating digital engineering.

The DSB report on test and evaluation also identified a need to modernize test infrastructure and recommended that DOD reevaluate test infrastructure to support future needs. It also recommended that the Test Resource Management Center—a group within OUSD(R&E) that manages test and evaluation infrastructure department-wide—accelerate the development of a knowledge management system and assess emerging commercial capabilities in autonomous vehicles and sensing that could be leveraged. DSB also recommended that OUSD(R&E) explore approaches to increase the priority of test and evaluation infrastructure in military construction prioritization discussions to identify funding to meet test and evaluation needs.

DSB found, and test officials similarly stated to us, that funding structures pose a challenge to modernizing test and evaluation. The DSB report on digital engineering found that the cost of acquiring tools and applications is a barrier for some programs to use digital engineering. For example, both Air Force and Navy officials told us that investing in digital tools or other test infrastructure can be a challenge for some programs that have less funding to invest in these tools, because testers rely on programs for funding rather than having funding of their own to invest in tools. The Test Resource Management Center provides funding to develop and acquire new test capabilities and address modernization projects too large for a single military department, but such projects must satisfy multi-service requirements.

Workforce Transformation

Plans from test organizations DOD-wide and at each military department call for identifying workforce needs or for filling skills gaps that affect the organization's ability to modernize test and evaluation through training or recruitment.

- **Identifying skill gaps.** The DOD Strategic Management Plan calls for annual assessments of DOD's science and technology workforce to support critical technology development. Additionally, the DOT&E Science and Technology Strategic Plan calls for developing internal expertise and creating a pipeline of incoming workers with necessary skill sets for test and evaluation.
- **Training for new technology and processes.** The DSB test and evaluation report recommended that OUSD(A&S) and OUSD(R&E) develop a comprehensive approach to developing the test and evaluation

workforce to enable it to adapt to rapidly advancing technologies. The report also recommended that OUSD(A&S) ensure program and test staff have been trained in the use of needed digital tools and environments. Modernization efforts require program or testing staff to make changes to existing processes to affect speed of development and testing; training prepares the workforce to execute new processes. Plans from DOT&E, OUSD(R&E), and each military department emphasize the need to ensure training opportunities to prepare the workforce for modernizing test and evaluation.

Officials at DOD-wide and military department test organizations told us that an impediment to modernization is that adherence to existing practices prevents change. For example, AFTC officials said that, historically, findings from test events were not typically shared until the test reports were signed and provided to programs. This created lag time between when issues were identified during testing and when those issues were addressed. These officials said that the cultural shift is needed to ensure that contractors and the test community are more collaborative. Some officials described actions taken by their organization to improve processes or enhance collaboration.

- **Hiring flexibilities.** Officials at multiple military departments told us that they have difficulty competing with the private sector in recruiting and retaining staff with necessary skill sets for digital transformation due to the differences in government pay scales and private sector compensation. As part of its fiscal year 2024 Campaign Plan, ATEC identified efforts to promote the use of incentives and direct hiring authorities to improve recruitment and stated an intent to implement programs aimed at supporting staff to improve work-life balance and workforce retention.

Policy and Guidance Revisions

Plans from DOT&E, OUSD(R&E), the Army, and the Navy call for updating policies and guidance to address the need to test new and emerging technologies or to leverage new tools that can be used in test and evaluation. DSB's recommendations call for a strategic shift in testing and point toward the need for additional policy or guidance.

- **Testing new technologies.** The DOT&E Strategy Implementation Plan states that updated guidance is needed to test space systems and AI-enabled systems, and to use advanced tools to support operational test activities.
- **Employing digital engineering.** OUSD(R&E)'s "Test and Evaluation as a Continuum" framework states that OSD and the military departments should develop and employ a common test and evaluation modeling system—including data dictionaries—and contracting guidance consistent with test and evaluation as a continuum.[19] The digital engineering strategies issued by OSD, the Army, and the Navy acknowledge digital engineering's role in facilitating a shift from the traditional method of capability development to a more flexible approach that allows for prototyping, experimentation, and testing in a virtual environment. These strategies call for developing policy and guidance related to digital engineering.

  The DSB task force on digital engineering found, among other things, that DOD lacks common standards in digitally representing environments and systems. It recommended that DOD develop common digital reference architecture and infrastructure to support the implementation of digital engineering across services and systems. Additionally, the DSB test and evaluation task force recommended that OUSD(R&E)

[19]Data dictionaries are documents that contain information on data elements (or fields), their definitions, descriptions, codes, and values.

publish best practices on the use of digital engineering and guidance on developmental testing requirements to enable efficient software development and use of digital engineering principles.

- **Testing early and continuously.** The DSB report on test and evaluation calls for a strategic shift in test and evaluation from the current acquisition-based framework to one that incorporates continuous development and testing. It made several recommendations aimed at this strategic shift that would likely require new policy and guidance to implement. For example, it recommended that:
  - OUSD(A&S), OUSD(R&E) and DOT&E expand their view of test and evaluation activities to incorporate verification and validation activities that occur before initiation of formal acquisition programs to accelerate the transition to the formal program by avoiding repetition;
  - OUSD(A&S) direct service acquisition executives—military department-level acquisition executives responsible for implementing DOD acquisition policy within their respective department—to structure new programs to incorporate testability requirements, maximize use of automation to increase testing, and develop approaches to report data that can be used to improve system performance; and
  - OUSD(R&E) develop developmental testing guidance to ensure system capability to use automated developmental testing to the maximum extent possible.

Collectively, these modernization plans identified a need for DOD to issue new or revised policies and guidance to support modernized test and evaluation.

Test and Evaluation Modernization Is Constrained by Policies That Do Not Fully Reflect Leading Practices

DOD-wide policies related to test and evaluation do not fully reflect relevant leading practices for product development. For example, the policies specific to test and evaluation largely did not reflect those leading practices for the major capability acquisition and middle tier of acquisition pathways, which account for the majority of DOD's spending on weapon systems acquisition. Instances where we found these leading practices take root were largely confined to the software acquisition and urgent capability acquisition pathways. Other related DOD-wide functional area policies on digital engineering and systems engineering largely did not include processes for how testers can apply these leading practices. Further, military department test and evaluation policies generally do not reflect leading practices beyond what is reflected in applicable DOD-wide policies. Correspondingly, in our review of acquisition strategies and test strategies for selected programs across the military departments, we found no evidence that these programs have operationalized relevant leading practices.

DOD-wide Policies Related to Test and Evaluation Do Not Fully Reflect Relevant Leading Practices for Product Development

DOD-wide policies related to test and evaluation do not fully reflect key tenets of our leading practices for product development.[20] The policies specific to test and evaluation generally reflect leading practices applicable to certain acquisition pathways, such as the software acquisition and urgent capability acquisition pathways.[21] In June 2025, we found that programs in two other acquisition pathways—major capability acquisition and middle-tier of acquisition—account for the largest spending on DOD's weapon system acquisitions.[22] However, DOD-wide test and evaluation policies did not fully reflect leading practices for those acquisition pathways. Further, we found that two DOD-wide functional area policies that support test and evaluation—DOD Instruction 5000.97 on digital engineering and DOD Instruction 5000.88 on systems engineering—reflect some elements of leading practices for product development. These policies, however, do not define or describe any specific processes for ensuring their application to test and evaluation (see fig. 4).

[20]Department of Defense Instruction 5000.89, *Test and Evaluation* (Nov. 19, 2020); Department of Defense Instruction 5000.98, *Operational Test and Evaluation and Live Fire Test and Evaluation* (Dec. 9, 2024); Department of Defense Instruction 5000.88, *Engineering of Defense Systems* (Nov. 18, 2020); Department of Defense Instruction 5000.97, *Digital Engineering* (Dec. 21, 2023).

In November 2020, DOD issued DOD Instruction 5000.89 as policy for both developmental and operational test and evaluation. In December 2024, DOD issued DOD Instruction 5000.98, which superseded the operational test and evaluation policy previously established in DOD Instruction 5000.89. DOD Instruction 5000.89 remains the current policy for developmental test and evaluation.

[21]Department of Defense Instruction 5000.89, *Test and Evaluation* (Nov. 19, 2020); Department of Defense Instruction 5000.98, *Operational Test and Evaluation and Live Fire Test and Evaluation* (Dec. 9, 2024).

[22]In June 2025, we reported that DOD planned to invest approximately $2.3 trillion across 79 major defense acquisition programs, which are included in the major capability acquisition pathway. DOD also planned to invest approximately $44.5 billion across 20 programs in the middle tier of acquisition pathway. GAO-25-107569.

Figure 4: Key DOD-wide Policies Applicable to Weapon System Test and Evaluation Do Not Fully Reflect Selected Leading Practices

Selected leading practices for product development		Application of leading practice to test and evaluation	Test and evaluation policies: DODI 5000.89 November 2020 (current developmental test policy) — Overall	Test and evaluation policies: DODI 5000.89 November 2020 (current developmental test policy) — MCA and MTA	Test and evaluation policies: DODI 5000.98 December 2024 (current operational and live-fire test policy) — Overall	Test and evaluation policies: DODI 5000.98 December 2024 (current operational and live-fire test policy) — MCA and MTA	Other test-related policies: DODI 5000.88 November 2020 (systems engineering policy) — Overall	Other test-related policies: DODI 5000.97 December 2023 (digital engineering policy) — Overall
Involvement in early acquisition decisions	Employ and empower right-sized teams of multi-disciplined stakeholders	Involve test officials early in defining the development program, including in the development of the acquisition strategy and by providing access to the key items, such as digital twins and digital threads, needed to conduct testing.	◐	MCA ○ / MTA ◐	◐	◐	○	○
Iterative development	Implement iterative testing processes to generate a minimum viable product	Develop test plans that reflect an iterative testing approach and integrate developmental and operational testing to support rapid development of a minimum viable product.	◐	○	◐	○	○	○
	Use modern design tools, including digital twins and digital threads	Ensure testers use key modern design tools, including digital twins and digital threads developed by the design contractor and needed to support iterative government testing.	◐	○	◐	◐	○	○
User feedback	Actively engage users to obtain feedback throughout an iterative development process	Develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input throughout testing to ensure the capability under development is relevant and responsive to users' most critical needs.	◐	○	◐	○	○	○

● Policy reflects leading practice ◐ Policy partially reflects leading practice ○ Policy does not reflect leading practice

DODI = Department of Defense (DOD) Instruction

MCA = Major capability acquisition
MTA = Middle tier of acquisition
} Majority of DOD spending on weapon systems acquisition

Source: GAO analysis of DOD-wide policies applicable to test and evaluation. | GAO-26-107009

Accessible Data for Figure 4: Key DOD-wide Policies Applicable to Weapon System Test and Evaluation Do Not Fully Reflect Selected Leading Practices

Test and evaluation policies	Test and evaluation policies	Test and evaluation policies	Test and evaluation policies	Other test-related policies	Other test-related policies
DODI 5000.89 November 2020 (current developmental test policy)	DODI 5000.89 November 2020 (current developmental test policy)	DODI 5000.98 December 2024 (current operational and live-fire test policy)	DODI 5000.98 December 2024 (current operational and live-fire test policy)	DODI 5000.88 November 2020 (systems engineering policy)	DODI 5000.97 December 2023 (digital engineering policy)
Overall	MCA and MTA	Overall	MCA and MTA	Overall	Overall
Partially reflects	MCA does not reflect; MTA partially reflects	Partially reflects	Partially reflects	Does not reflect	Does not reflect
Partially reflects	Does not reflect	Partially reflects	Does not reflect	Does not reflect	Does not reflect
Partially reflects	Does not reflect	Partially reflects	Partially reflects	Does not reflect	Does not reflect
Partially reflects	Does not reflect	Partially reflects	Does not reflect	Does not reflect	Does not reflect

Source: GAO analysis of DOD-wide policies applicable to test and evaluation. | GAO-26-107009

Notes: The "overall" determinations reflect GAO's assessments of the full contents of a given policy applicable to the four weapon system acquisition pathways: major capability acquisition, middle tier of acquisition, urgent capability acquisition, and software acquisition. For the test-specific and test-related policies, GAO's policy determinations for partially reflects indicate at least one of the following: (1) the policy language under review reflects some, but not all, parts of GAO's selected leading practice description as it applies to test and evaluation; or (2) the policy language fully reflects GAO's selected leading practice but only for some of the acquisition pathways directly related to weapon systems in the Adaptive Acquisition Framework. For the test-specific policies, a partial determination for the major capability acquisition and middle tier of acquisition pathways indicates that the policy partially reflects the relevant leading practice for at least one of the pathways.

DOD Instruction 5000.98 establishes policy, assigns responsibilities, and prescribes procedures for operational test and evaluation of DOD systems and services acquired via the Defense Acquisition System or via other non-standard acquisition systems. DOD Instruction 5000.98 supersedes information regarding operational test and evaluation located in DOD Instruction 5000.89.

- **Involvement in early acquisition decisions.** DOD-wide test and evaluation policies state that programs should set up a test and evaluation working-level integrated product team, which we refer to as a cross-functional test team. The cross-functional test team assists in developing the test strategy and is the primary forum for collaboration between the program manager and test community. These test and evaluation policies also state the programs should set up the cross-functional test team early in the program, but the policies are not specific as to when.

  Specifically, DOD Instruction 5000.89—herein referred to as the DOD-wide developmental test and evaluation policy—includes tester involvement in the development of the acquisition strategy, including the development of testing requirements, for the software acquisition and middle tier of acquisition pathways.[23] However, there is no similar language for the other acquisition pathways, including the major capability acquisition pathway. The policy also does not state whether testers are to have access to digital twins or digital threads, which are key items needed to conduct testing in an iterative development framework. OUSD(R&E) drafted revisions to its developmental test and evaluation policy that, if implemented, would include provisions placing greater emphasis on early tester involvement. For example, one draft provision would require the involvement of the cross-functional test team in the development of requirements and other critical acquisition documents including the acquisition strategy and the contract.

[23]Department of Defense Instruction 5000.89, *Test and Evaluation* (Nov. 19, 2020). For middle tier of acquisition programs, the test strategy—which is included as part of the acquisition strategy—should detail required test resources, including test tools.

Similarly, DOD Instruction 5000.98—herein referred to as the DOD-wide operational test and evaluation policy—states that testers should be involved in the development of the acquisition strategy and the inclusion of test requirements for programs in all pathways, including the major capability acquisition and middle tier of acquisition pathways.[24] However, the test policy does not identify whether the testers have access to digital twins or digital threads.

DOD-wide digital engineering policy further states that programs generally should incorporate digital engineering for the capability in development, and that digital engineering must be addressed in the acquisition strategy, including how and when it will be used in the system life cycle. However, the policy does not further define a process by which testers can advocate for the access to digital twins and digital threads during development of the acquisition strategy. Similarly, the systems engineering policy does not provide that additional context.

- **Iterative development.** DOD-wide developmental test and evaluation policy requires programs in the software acquisition pathway to develop test strategies that detail how they plan to use iterative test processes. However, it does not require test strategies or test plans to address iterative testing for acquisition programs in other pathways, including the major capability acquisition and middle tier of acquisition pathways.[25] Additionally, the policy requires all DOD acquisition programs, regardless of acquisition pathway, to execute iterative developmental and operational testing for cybersecurity throughout the program's life cycle. The policy also requires testers to consider using digital twins for some programs in the software acquisition pathway but does not require testers to use digital twins for other pathways, including major capability and middle tier of acquisition pathways. OUSD(R&E) drafted revisions to its developmental test and evaluation policy that, if implemented, would emphasize a shift away from the traditional testing approach for acquisition programs toward an iterative approach to support an agile decision-making process. For example, one draft provision would require that, in this new approach, developmental testing encompass a campaign of learning covering the entire capability life cycle, which would provide focused and relevant information to support decision-making with the goal of accelerating the fielding of capabilities. This provision, if implemented, would reflect OUSD(R&E)'s "Test and Evaluation as a Continuum" framework for modernizing test and evaluation.

  Similarly, DOD-wide operational test and evaluation policy requires test plans for programs in the software acquisition pathway to use an iterative test process but does not require test strategies or test plans to address iterative testing for acquisition programs in other pathways.[26] The policy also references the use of digital engineering and digital tools such as digital twins, though it does not reference the use of digital threads.

  DOD-wide digital engineering policy states that OUSD(R&E) is responsible for implementing digital engineering in the activities for developmental testing, and states that DOT&E supports the use of digital engineering in operational testing. It also states that DOD will develop a digital engineering capability that includes digital twins and threads. However, it does not describe how the test community will use those digital twins and threads in their test processes, including the development of test strategies and test plans. Similarly, the DOD-wide systems engineering policy does not expand upon those areas.

- **User feedback.** DOD-wide test and evaluation policies generally do not require testers to iteratively engage users to obtain feedback throughout testing to support system development. Specifically, DOD-

[24]Department of Defense Instruction 5000.98, *Operational Test and Evaluation and Live Fire Test and Evaluation* (Dec. 9, 2024).

[25]Department of Defense Instruction 5000.89, *Test and Evaluation* (Nov. 19, 2020).

[26]Department of Defense Instruction 5000.98, *Operational Test and Evaluation and Live Fire Test and Evaluation* (Dec. 9, 2024).

wide developmental test and evaluation policy states that testing should generally include user feedback to support design and operational use improvements for the urgent capability acquisition pathway, and the DOD-wide operational test and evaluation policy reiterates this for the urgent capability acquisition pathway. However, there is no similar language for the other acquisition pathways, including the major capability acquisition and middle tier of acquisition pathways. The policies also do not specify how end user feedback is to be solicited routinely throughout iterative testing. Further, DOD-wide digital engineering and systems engineering policies do not expand on how the test community should solicit user feedback through iterative testing.

DOD acquisition policy requires programs using the software pathway to create user agreements.[27] User agreements are a commitment between the sponsor and program manager for continuous user involvement. The agreements assign decision-making authority for decisions about required capabilities, user acceptances, and readiness for operational deployment. These agreements help to ensure the user community is represented and engaged throughout development by defining responsibilities and expectations for involvement and interaction of users and developers. In July 2023, we recommended, among other things, that DOD incorporate agile principles into requirements policy and guidance for all programs using agile for software development, including a user agreement.[28] DOD partially concurred with this recommendation, and agreed to clarify requirements policy to provide guidance on using an acquisition Capability Needs Statement and user agreement for development of software that is embedded within an already validated requirements document without needing additional requirements validation. This recommendation remains open as of May 2025.

Without revising DOD-wide test and evaluation policies and related functional area policies to fully reflect the relevant leading practices for all acquisition pathways directly related to weapon systems, DOD is missing opportunities to ensure that the systems under test are responsive to and relevant to warfighter needs. Such practices include obtaining early input from testers on the acquisition strategy, reflecting an iterative test approach in test strategies and test plans, and developing user agreements in test strategies and test plans to obtain user feedback.

Military Department Test and Evaluation Policies Reflect Marginal Additional Use of Leading Practices Beyond DOD-Wide Policies

Military department test and evaluation policies reflect relevant leading practices to only a limited extent beyond what is in the DOD-wide policies.[29] Further, we found no evidence in our review of program documentation that DOD has operationalized those leading practices in its test and evaluation function. Figure

[27]Department of Defense Instruction 5000.87, *Operation of the Software Acquisition Pathway* (Oct. 2, 2020).

[28]GAO, *Defense Software Acquisitions: Changes to Requirements, Oversight, and Tools Needed for Weapon Programs*, GAO-23-105867 (Washington, D.C.: July 20, 2023).

[29]The military departments are required to adhere to the requirements in the DOD-wide policies. The role of the military department test and evaluation policies is to implement DOD-wide policies within the military department's specific context. Accordingly, we analyzed the military department policies to see if their implementation of DOD-wide policy reflected the leading practices beyond the level established in the DOD-wide policies. In most cases, the military department test and evaluation policies do not reflect the leading practices beyond the limited application we found in the DOD-wide policies, including for the major capability acquisition and middle tier of acquisition pathways. In another case, we found that one of the military department test and evaluation policies repeated some tenets of the leading practices that are already captured in the DOD-wide policies. In other cases, we found that military department policies did not repeat requirements detailed in the DOD-wide policies. Under our methodology, we did not credit the military departments for any language already reflected as part of our analysis of the DOD-wide policies.

5 summarizes the extent to which military department test and evaluation policies reflect our selected leading practices for product development beyond what is already reflected in DOD-wide policies.

Figure 5: Military Department Weapon System Test and Evaluation Policies Do Not Fully Reflect Selected Leading Practices Beyond the Level in DOD-Wide Policy

Selected leading practices for product development		Application of leading practice to test and evaluation	Test and evaluation policies: Air Force Instruction 99-103 — Overall	Air Force — MCA and MTA	Army Regulation 73-1 — Overall	Army — MCA and MTA	Navy Instruction 5000.2G, Section 10: Test and Evaluation — Overall	Navy — MCA and MTA
Involvement in early acquisition decisions	Employ and empower right-sized teams of multi-disciplined stakeholders	Involve test officials early in defining the development program, including in the development of the acquisition strategy and by providing access to the key items, such as digital twins and digital threads, needed to conduct testing.	◐	MCA ◐ / MTA ○	◐	Not applicable	○	○
Iterative development	Implement iterative testing processes to generate a minimum viable product	Develop test plans that reflect an iterative testing approach and integrate developmental and operational testing to support rapid development of a minimum viable product.	○	○	○		○	○
	Use modern design tools, including digital twins and digital threads	Ensure testers use key modern design tools, including digital twins and digital threads developed by the design contractor and needed to support iterative government testing.	○	○	○		○	○
User feedback	Actively engage users to obtain feedback throughout an iterative development process	Develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input throughout testing to ensure the capability under development is relevant and responsive to users' most critical needs.	○	○	○		○	○

● Policy reflects leading practice ◐ Policy partially reflects leading practice ○ Policy does not reflect leading practice

MCA = Major capability acquisition
MTA = Middle tier of acquisition
} Majority of Department of Defense spending on weapon systems acquisition

Source: GAO analysis of military department test and evaluation policies. | GAO-26-107009

Notes: The "overall" determinations reflect GAO's assessments of the full contents of a given policy applicable to the four weapon system acquisition pathways: major capability acquisition, middle tier of acquisition, urgent capability acquisition, and software acquisition. GAO assigned ratings for the military department test and evaluation policies by determining whether the policy language further reflected leading practices for product development beyond the policy established in DOD Instruction 5000.89 and DOD Instruction 5000.98. Air Force, Army, and Navy test and evaluation policies incorporate and implement DOD-wide test and evaluation policy. See figure 4 for GAO's analysis of DOD-wide test and evaluation policy.

A partial rating indicates at least one of the following: (1) the policy language under review reflects some, but not all, parts of GAO's selected leading practice for product development as it applies to test and evaluation; or (2) the policy language fully reflects GAO's selected leading practice but for some, but not all, acquisition pathways directly related to weapon systems in the Adaptive Acquisition Framework.

The Army issued its test and evaluation policy in June 2018, prior to the establishment of the Adaptive Acquisition Framework in 2020, and as such, does not contain any acquisition pathway specific test policy or procedures. Therefore, GAO did not evaluate the policy against specific pathways.

Air Force Test and Evaluation Policy Does Not Fully Reflect Leading Practices

Air Force test and evaluation policy partially reflects one leading practice beyond what is in DOD-wide test and evaluation policies.[30] Our assessment also found the following:

- **Involvement in early acquisition decisions.** The current Air Force test and evaluation policy partially reflects the leading practice—tester involvement in the development of acquisition strategies and advocating for access to digital twins and digital threads—beyond the DOD-wide test and evaluation policies.[31] Specifically, the Air Force policy requires developmental testers to be involved in the development of program acquisition strategies irrespective of the selected acquisition pathway. This is an improvement over DOD-wide policies, which include developmental tester involvement for only those programs on the software acquisition and middle tier of acquisition pathways. Like the DOD-wide policies though, the Air Force policy does not require testers to advocate for access to digital twins and digital threads—a shortfall that is inconsistent with the leading practice.

  Air Force test officials are not always involved early in the acquisition process. Air Force developmental test officials said that a cross-functional test team is usually established too late in the acquisition process. Those officials added that Air Force testers rely on acquisition officials to acquire program models and data through the contract, and in some instances, it has taken an extra year of negotiations to get the data and models from the contractor. To illustrate this issue, figure 6 below provides information on tester involvement in developing the T-7A acquisition strategy.

[30]The Air Force's test and evaluation policy reprints the DOD developmental test and evaluation policy word-for-word without editorial review. The policy separately indicates which portions are Department of the Air Force Instruction supplementary material. Our analysis of the Air Force test and evaluation policy only analyzes the Department of the Air Force Instruction supplementary material and not the reprinting of DOD-wide policy. Department of the Air Force Instruction 99-103, *Capabilities-Based Test and Evaluation* (Dec. 9, 2021) (incorporating change, Feb. 28, 2025).

[31]Department of the Air Force Instruction 99-103, *Capabilities-Based Test and Evaluation* (Dec. 9, 2021) (incorporating change, Feb. 28, 2025).

Figure 6: T-7A Acquisition Strategy Does Not Describe Timely Access to Contractor Test Data and Digital Model



T-7A Red Hawk

Development start: September 2018

Operational testing start: June 2026 (almost 8 years after development start)

Initial operational capability: January 2028

Cycle time: about 9 years

The Air Force's T-7A Red Hawk program is expected to replace the Air Force's legacy T-38C trainer fleet and related ground equipment and address the Air Force's advanced fighter pilot training needs.

Air Force test officials said that they did not have timely access to T-7A test data. Specifically, they said there were challenges with the transfer of large data files due to cybersecurity requirements complicating physical connection of government and contractor proprietary high-speed data networks. This resulted in government personnel hand-carrying hard drives to Edwards Air Force Base until a contractor terminal could be installed there for downloading terabytes of historical test data. After the test aircraft were delivered, data could be downloaded directly from aircraft to government networks.

An Air Force developmental tester said that the acquisition strategy was developed based on input of developmental and operational testers. However, the acquisition strategy for the Air Force's T-7A program does not discuss obtaining digital access to timely contractor test data that testers could leverage to inform their own testing and assessments for product development as part of an iterative development cycle. Further, Air Force test officials said that the contract awarded did not require a complete digital system model from the prime contractor.

Source: GAO review of Department of Defense information; Boeing Corporation (image). | GAO-26-107009

Accessible Data for Figure 6: T-7A Acquisition Strategy Does Not Describe Timely Access to Contractor Test Data and Digital Model

T-7A Red Hawk

Development start: September 2018

Operational testing start: June 2026 (almost 8 years after development start)

Initial operational capability: January 2028

Cycle time: about 9 years

The Air Force's T-7A Red Hawk program is expected to replace the Air Force's legacy T-38C trainer fleet and related ground equipment and address the Air Force's advanced fighter pilot training needs.

Air Force test officials said that they did not have timely access to T-7A test data. Specifically, they said there were challenges with transfer of large data files due to cybersecurity requirements complicating physical connection of government and contractor proprietary high-speed data networks. This resulted in government personnel hand-carrying hard drives to Edwards Air Force Base until a contractor terminal could be installed there for downloading terabytes of historical test data. After the test aircraft were delivered, data could be downloaded directly from aircraft to government networks.

An Air Force developmental tester said that the acquisition strategy was developed based on input of developmental and operational testers. However, the acquisition strategy for the Air Force's T-7A program does not discuss obtaining digital access to timely contractor test data that testers could leverage to inform their own testing and assessments for product development as part of an iterative development cycle. Further, Air Force test officials said that the contract awarded did not require a complete digital system model from the prime contractor.

Source: GAO review of Department of Defense and Air Force information; Boeing Corporation (Image). | GAO-25-107009

- **Iterative development.** Air Force test and evaluation policy does not reflect the leading practices—developing test strategies and test plans that reflect an iterative test approach and use digital twins—beyond the DOD-wide test and evaluation policies. The Air Force drafted revisions to its test and evaluation policy that, if implemented, would include requirements for testers to evaluate digital models of systems, digital twins, and digital threads as components of the supporting infrastructure for test analysis and evaluation planning across all Adaptive Acquisition Framework pathways.

  Additionally, none of the acquisition strategies and test plans associated with the three Air Force programs we reviewed considered the use of digital twins and digital threads for test and evaluation to support iterative product development.

- **User feedback.** Air Force test and evaluation policy does not reflect the leading practice—developing test strategies and test plans that detail a process for obtaining ongoing user feedback—beyond the DOD-wide test and evaluation policies.

  For the selected Air Force programs we reviewed, the test plans did not identify how ongoing user feedback can be used to support an iterative test process. As a result, programs across pathways are not ensuring that the weapon system under test meets essential user needs. Figure 7 below provides information about the extent to which the F-22 program incorporated end user feedback into its test plans.

Figure 7: Test Plans for Air Force's F-22 SeE Program Do Not Consider How Ongoing User Feedback Could Be Incorporated into an Iterative Test Process



F-22 Sensor Enhancements (F-22 SeE)

Program start: September 2018 (middle tier of acquisition (MTA) rapid prototyping initiation), December 2018 (sensor enhancements added to program); August 2024 (F-22 SeE MTA rapid fielding initiation)

First flight test (Program has a planned date for first flight test, but does not have an estimated date for operational testing): January to March 2026 (1.5 years after the rapid prototyping program transitioned into a rapid fielding program, and about 7 years after the sensor enhancements were added to the rapid prototyping program)

Initial operational capability: January to March 2029

Cycle time: about 4 years after the rapid prototyping program transitioned into a rapid fielding program, and about 10 years after sensor enhancements were added to the rapid prototyping program

The F-22 program—previously utilizing the MTA rapid prototyping pathway and now utilizing the rapid fielding pathway—intends to develop, integrate, and deliver hardware and software capabilities, including for tactical information transmission, combat identification, navigation, sensors, fuel tanks, and electronic protection.

The two F-22 test plans that GAO reviewed consider the collection of user input but do not detail how this feedback can be collected on an ongoing basis to inform an iterative test process. Air Force test officials said that there is a working group that collects user feedback from F-22 combat pilots, and that user feedback is collected using operational test aircraft that receive software updates within days of the developmental test aircraft. However, test officials did not describe a process of continuous user engagement to collect feedback on system design that would inform product development in an iterative test approach.

Source: GAO review of Department of Defense and Air Force information; Defense Visual Information Distribution Service (image). | GAO-26-107009

Accessible Data for Figure 7: Test Plans for Air Force's F-22 SeE Program Do Not Consider How Ongoing User Feedback Could Be Incorporated into an Iterative Test Process

F-22 Sensor Enhancements (F-22 SeE)

Program start: September 2018 (middle tier of acquisition (MTA) rapid prototyping initiation), December 2018 (sensor enhancements added to program); August 2024 (F-22 SeE MTA rapid fielding initiation)

First flight test*: January to March 2026 (1.5 years after the rapid prototyping program transitioned into a rapid fielding program, and about 7 years after the sensor enhancements were added to the rapid prototyping program)

Initial capability: January to March 2029

Cycle time: about 4 years after the rapid prototyping program transitioned into a rapid fielding program, and about 10 years after sensor enhancements were added to the rapid prototyping program

*Program has a planned date for first flight test, but does not have an estimated date for operational testing.

The F-22 program—previously utilizing the MTA rapid prototyping pathway and now utilizing the rapid fielding pathway—intends to develop, integrate, and deliver hardware and software capabilities, including for tactical information transmission, combat identification, navigation, sensors, fuel tanks, and electronic protection.

The two F-22 test plans that GAO reviewed consider the collection of user input but do not detail how this feedback can be collected on an ongoing-basis to inform an iterative test process. Air Force test officials said that there is a working group that collects user feedback from F-22 combat pilots, and that user feedback is collected using operational test aircraft that receive software updates within days of the developmental test aircraft. However, test officials did not describe a process of continuous user engagement to collect feedback on system design that would inform product development in an iterative test approach.

Source: GAO review of Department of Defense and Air Force information; Defense Visual Information Distribution Service (Image). | GAO-25-107009

Tester involvement in the development of the acquisition strategy could better enable test officials to get access to contractor test data. This would allow the test community to use leading practices for product development and conduct systems-integrated tests on a digital twin or a physical prototype connected to a digital twin to assess whether the minimum viable product meets user needs.

Without revising Air Force test and evaluation policy to fully reflect the relevant leading practices for all acquisition pathways directly related to weapon systems, the Air Force is missing opportunities to ensure that testing proceeds expeditiously and that the systems under test are responsive and relevant to warfighter needs. Such practices include advocating for tester access to digital twins and digital threads as part of acquisition strategy development, reflecting an iterative test approach in test strategies and test plans, and developing user agreements in test strategies and test plans to obtain user feedback.

Army Test and Evaluation Policy Does Not Fully Reflect Leading Practices

The Army's test and evaluation policy reflects leading practices to a limited extent beyond what is in DOD-wide test and evaluation policies. Specifically, the current policy states that testers should have early involvement in the acquisition strategy. The Army's test and evaluation policy predates the Adaptive Acquisition Framework that established the different acquisition pathways. As such, this policy provides a common set of test and evaluation policies irrespective of the different pathways. Our assessment of the policy found the following:

- **Involvement in early acquisition decisions.** The Army test and evaluation policy requires tester involvement in the development of acquisition strategies for all programs—whereas the DOD-wide developmental test policy requires it only for the software acquisition and middle tier of acquisition pathways.[32] However, similar to the DOD-wide test and evaluation policy, the Army test and evaluation policy does not require testers to advocate for access to digital twins and digital threads.

  Army test officials said they are not always involved early in the acquisition process, and therefore not proactively involved in getting test data requirements into the contract. Instead, Army test officials said they must work through the program manager to get access to data. Those officials added that they are working to be more involved earlier in the acquisition process, but they believe that programs are beginning to embrace testers as part of the team developing the weapon system.

- **Iterative development.** Army test and evaluation policy does not reflect the leading practices—developing test strategies and test plans that reflect an iterative test approach and use digital twins and digital threads—beyond the DOD-wide test and evaluation policies.

  The Army drafted revisions to its test and evaluation policy that, if implemented, would include provisions referencing the use of models and data to conduct integrated testing and to support an agile approach with continuous operational testing. The draft policy also would contain provisions requiring models and data to be used to digitally represent the system in a mission context to conduct integrated test and evaluation activities, and to the largest extent possible, programs would use an accessible digital ecosystem as part of the Army's digital engineering strategy. Army officials said they do not have an estimate for when they plan to publish the revised test policy.

  The Army acquisition strategies and test plans we reviewed generally did not advocate for government access to contractor-developed digital twins and digital threads and provided limited tester access to contractor data to inform product development. Additionally, the Army test plans that we reviewed inconsistently supported implementing an iterative government testing approach. Further, none of the selected Army acquisition strategies and test plans we reviewed considered the use of digital twins and digital threads for test and evaluation to support iterative product development. Figure 8 describes how the Future Long Range Assault Aircraft program intends to use digital tools, but not in an iterative test approach.

[32]The Army issued its test and evaluation policy in June 2018, prior to the establishment of the Adaptive Acquisition Framework in 2020, and does not contain any acquisition pathway specific test policy or procedures. Therefore, we did not evaluate the policy against specific pathways.

Figure 8: Army's FLRAA Program Does Not Have an Iterative Test Approach



Future Long Range Assault Aircraft (FLRAA)

Program start: October 2020 (middle tier of acquisition program initiation); July 2024 (transition to the major capability acquisition pathway)

Operational testing start: planned for fiscal year 2031 (about 7 years after initiating the major capability acquisition, and 11 years after the middle tier acquisition program)

Initial capability: planned for October to December 2032

Cycle time: about 8 years after program start for major capability acquisition, and 12 years after program start of the middle tier of acquisition program

FLRAA is part of the Future Vertical Lift portfolio of systems that is intended to be a medium-sized assault and utility aircraft and deliver speed, range, agility, endurance, and sustainability improvements.

While the Army's FLRAA program is using a digital thread, the program's test plan does not describe if and how testers plan to use the digital thread for test and evaluation. An official in the Office of the Director for Operational Test and Evaluation said the FLRAA test plan supports iterative testing for cybersecurity and live-fire testing; however, the test plan does not describe the use of iterative testing for the entire system.

Source: GAO review of Department of Defense and Army information; Bell Textron, Inc (image). | GAO-26-107009

Accessible Data for Figure 8: Army's FLRAA Program Does Not Have an Iterative Test Approach

Future Long Range Assault Aircraft (FLRAA)

Program start: October 2020 (middle tier of acquisition program initiation); July 2024 (transition to the major capability acquisition pathway)

Operational testing start: planned for fiscal year 2031 (about 7 years after initiating the major capability acquisition, and 11 years after the middle tier acquisition program)

Initial capability: planned for October to December 2032

Cycle time: about 8 years after program start for major capability acquisition, and 12 years after program start of the middle tier of acquisition program

FLRAA is part of the Future Vertical Lift portfolio of systems that is intended to be a medium-sized assault and utility aircraft and deliver speed, range, agility, endurance, and sustainability improvements.

While the Army's FLRAA program is using a digital thread, the program's test plan does not describe if and how testers plan to use the digital thread for test and evaluation. An official in the Office of the Director for Operational Test and Evaluation said the FLRAA test plan supports iterative testing for cybersecurity and live-fire testing; however, the test plan does not describe the use of iterative testing for the entire system.

Source: GAO review of Department of Defense and Army information; Bell Textron, Inc (Image). | GAO-25-107009

- **User feedback.** Army test and evaluation policy does not reflect the leading practice—developing test strategies and test plans that detail a process for obtaining ongoing user feedback—beyond the DOD-wide test and evaluation policies.

Without revising Army test and evaluation policy to fully reflect the leading practices for all acquisition pathways directly related to weapon systems, the Army is missing opportunities to provide testing that is tailored and responsive to rapid design iterations, and to ensure that integrated testing proceeds expeditiously to get new

capabilities to warfighters sooner. Such practices include obtaining early input from testers on the acquisition strategy, reflecting an iterative test approach in test strategies and test plans, and developing user agreements in test strategies and test plans to obtain user feedback.

Navy Test and Evaluation Policy Does Not Fully Reflect Leading Practices

The Navy's test and evaluation policy does not further reflect leading practices beyond what is in DOD-wide test and evaluation policies.[33] Our assessment found the following:

- **Involvement in early acquisition decisions.** Navy test and evaluation policy does not reflect the leading practice—tester involvement in the development of acquisition strategies and advocating for access to digital twins and digital threads—beyond what is in DOD-wide test and evaluation policies for all the Adaptive Acquisition Framework pathways.

  Navy operational test officials said that they are generally not involved in the development of acquisition strategies. For example, Navy operational test officials said that most programs are proactive in discussions with their office, but others delay coordinating with test officials because the program is working through technical challenges. Additionally, these test officials said that the current acquisition framework does not incentivize program managers to reach out to testers early because many programs choose to delay testing while addressing other issues.
- **Iterative development.** Navy test and evaluation policy does not reflect the leading practices—developing test strategies and test plans that reflect an iterative test approach and use digital twins and digital threads—beyond the DOD-wide test and evaluation policies.

  The Navy test plans that we reviewed did not consistently support implementing an iterative government testing approach. Additionally, the Navy acquisition strategies and test plans we reviewed did not consider the use of digital twins and digital threads for test and evaluation to support iterative product development. Furthermore, none of the Navy test plans identified how user feedback can be used to support an iterative test process. As a result, these programs are not positioned to ensure, on an ongoing basis throughout development, that the weapon systems under test meet essential user needs. Figure 9 identifies how the Navy approached iterative development for the Ford class aircraft carrier program.

[33]We assessed the Test and Evaluation section of the Secretary of the Navy Instruction 5000.2G, which Navy test and evaluation officials stated governs their efforts.

Figure 9: The Navy's Ford Class Aircraft Carrier Program Did Not Fully Implement an Iterative Test Process



CVN 78 *Gerald R. Ford* Nuclear Aircraft Carrier (CVN 78)

Program start: June 2000

Initial operational capability: December 2021

Operational testing start: August 2022 (about 22 years after program start)

Cycle time: about 21.5 years

The Navy developed the CVN 78 (or *Ford* class) nuclear-powered aircraft carrier to introduce new propulsion, aircraft launch and recovery, and survivability capabilities to the carrier fleet.

The Navy's CVN 78 *Ford* class carrier program test plan reflects an iterative test approach, but only for cybersecurity developmental testing. CVN 78 program officials stated that this test plan reflects the non-iterative development approaches that the Navy employed for the majority of CVN 78 subsystems. Correspondingly, the *Ford* class does not incorporate digital twins or a digital thread.

Source: GAO review of Navy information; U.S. Navy (Image). | GAO-26-107009

Accessible Data for Figure 9: The Navy's Ford Class Aircraft Carrier Program Did Not Fully Implement an Iterative Test Process

CVN 78 Gerald R. Ford Nuclear Aircraft Carrier (CVN 78)

Program start: June 2000

Initial operational capability: December 2021

Operational testing start: August 2022 (about 22 years after program start)

Cycle time: about 21.5 years

The Navy developed the CVN 78 (or Ford class) nuclear-powered aircraft carrier to introduce new propulsion, aircraft launch and recovery, and survivability capabilities to the carrier fleet.

The Navy's CVN 78 Ford class carrier program test plan reflects an iterative test approach, but only for cybersecurity developmental testing. The Navy's CVN 78 Ford Class carrier program test plan reflects an iterative test approach, but only for cybersecurity developmental testing. CVN 78 program officials stated that this test plan reflects the non-iterative development approaches that the Navy employed for the majority of CVN 78 subsystems. Correspondingly, the Ford class does not incorporate digital twins or a digital thread.

Source: GAO review of Navy information; U.S. Navy (Image). | GAO-25-107009

- **User feedback.** Navy test and evaluation policy does not reflect the leading practice—developing test strategies and test plans that detail a process for obtaining ongoing user feedback—beyond the DOD-wide test and evaluation policies.

  Navy operational test officials said that user feedback is often derived from different types of operational testing. However, we found that the Navy test plans we reviewed did not identify how user feedback can be used to support an iterative test process. As a result, programs across pathways are not ensuring that the weapon system under test meets essential user needs.

Without revising Navy test and evaluation policy to fully reflecting the leading practices for all acquisition pathways directly related to weapon systems, the Navy is missing opportunities to ensure that the systems under test are responsive and relevant to warfighter needs. Such practices include obtaining early input from

testers on the acquisition strategy, reflecting an iterative test approach in test strategies and test plans, and developing user agreements in test strategies and test plans to obtain user feedback

Fully Realizing Test and Evaluation Modernization Hinges on Changes to Acquisition Strategy Development Policy

Multiple organizations are responsible for ensuring the test community is involved to effectively develop weapon systems. OUSD(A&S) is responsible for issuing acquisition policies to ensure an effective partnership between the program and test community. OUSD(A&S) also has a role in ensuring that the test community is consulted early and that test official input is obtained and incorporated into the acquisition strategy. In March 2022, we recommended that OUSD(A&S) update its acquisition policies to fully implement our leading practices throughout development, including obtaining a sound business case.[34] As part of a sound business case, programs should ensure they have stakeholder—anyone with interest in the product—input early in the program. DOD concurred with our recommendations, but as of February 2025 had yet to implement them. DOD is currently revising some of its acquisition policies that may yet introduce these practices.[35]

Test officials have a vested interest in weapon system acquisitions because they ultimately assess whether the systems are deemed suitable, effective, and lethal for the warfighter. Acquisition strategies set the program manager's plan to achieve program execution and programmatic goals across a program's entire life cycle and provide a basis for more detailed planning. Test officials provide input into the key items—such as digital twins and digital threads, and other resources—that are needed to test weapon systems and rapidly deliver capabilities.

As discussed, DOD acquisition programs have missed opportunities to obtain and incorporate test official input into the acquisition strategy. We found that, generally, cross-functional test teams are available to support acquisition programs. However, acquisition programs are not always using cross-functional test teams to help develop the acquisition strategy. Specifically, program officials for nearly half the programs told us that they did not seek input from the test and evaluation community while developing the programs' acquisition strategies. Further, we found that some of the acquisition programs did not form their cross-functional test teams until after the acquisition strategy was developed. As we previously discussed, the acquisition strategies for these programs generally did not include information on key items, such as digital threads and digital twins, needed to conduct testing.

OUSD(A&S) officials said that they are currently updating their acquisition policies to identify when the test community should be involved in the development of the acquisition strategy. However, because these policy revisions were still in development at the time of our review, we were unable to assess how the revisions addressed access to digital engineering tools. By not requiring early tester involvement in the development of the acquisition strategy in both acquisition and test and evaluation policy, DOD is missing opportunities to optimize acquisition strategies and test more efficiently to avoid the typical problems that reveal themselves downstream in weapon system acquisition programs.

[34]GAO-22-104513.
[35]GAO-25-107743.

Conclusions

Through its ongoing modernization planning, DOD has a unique opportunity to not only retool its existing test and evaluation enterprise, but to redefine the role that enterprise can play in enabling faster delivery of relevant capabilities to warfighters. Our prior work identifying the key practices that leading companies employ to successfully develop new products can guide DOD's efforts. However, modernization plans that advocate rapid, iterative test and evaluation approaches—consistent with leading practices—face headwinds from existing policies that predominantly continue to reflect traditional, linear approaches. These policies, which include ones issued by the Office of the Secretary of Defense, Air Force, Army, and Navy, could hamstring potential gains in speed that test and evaluation modernization might achieve.

DOD therefore faces a choice. On the one hand, it could continue to treat test and evaluation as a means of identifying deficiencies in weapon systems already developed and, often, already produced. This would require no changes to existing policy and the status quo, but it does not further DOD's goal of accelerating delivery of relevant capabilities to warfighters. On the other hand, DOD has an opportunity to embrace test and evaluation as a core foundation of every weapon systems acquisition program. Specifically, it could organize weapon system acquisition strategies, test plans, and strategies around iterative test and evaluation processes, enabled by digital twins and digital threads, that center on rapidly developing minimum viable products. By revising their test and evaluation policies to reflect such leading practices, DOD and the military departments would structure test and evaluation in ways that proactively inform and aid development of weapon systems' performance requirements and design, underpinned by ongoing and persistent user feedback. This approach, while potentially a significant change in the short term, over the long term would leave DOD better positioned to ensure that warfighter needs are consistently understood, prioritized, and met.

Recommendations for Executive Action

We are making a total of 13 recommendations, including four to the Secretary of Defense (Recommendations 1-3 and 13), three to the Secretary of the Air Force (Recommendations 4-6), three to the Secretary of the Army (Recommendations 7-9), and three to the Secretary of the Navy (Recommendations 10-12):

The Secretary of Defense should ensure that the Under Secretary of Defense for Research and Engineering and the Director of Operational Test and Evaluation revise their weapon system test and evaluation, digital engineering, and systems engineering policies to fully reflect leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads. (Recommendation 1)

The Secretary of Defense should ensure that the Under Secretary of Defense for Research and Engineering and the Director of Operational Test and Evaluation revise their weapon system test and evaluation, digital engineering, and systems engineering policies to fully reflect leading practices for product development by requiring developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products. (Recommendation 2)

The Secretary of Defense should ensure that the Under Secretary of Defense for Research and Engineering and the Director of Operational Test and Evaluation revise their weapon system test and evaluation, digital engineering, and systems engineering policies to fully reflect leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test. (Recommendation 3)

The Secretary of the Air Force should, as the Air Force revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads. (Recommendation 4)

The Secretary of the Air Force should, as the Air Force revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products. (Recommendation 5)

The Secretary of the Air Force should, as the Air Force revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test. (Recommendation 6)

The Secretary of the Army should, as the Army revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads. (Recommendation 7)

The Secretary of the Army should, as the Army revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products. (Recommendation 8)

The Secretary of the Army should, as the Army revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test. (Recommendation 9)

The Secretary of the Navy should revise the military department’s weapon system test and evaluation policy to fully reflect leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads. (Recommendation 10)

The Secretary of the Navy should revise the military department’s weapon system test and evaluation policy to fully reflect leading practices for product development by requiring developmental and operational testers to

develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products. (Recommendation 11)

The Secretary of the Navy should revise the military department's weapon system test and evaluation policy to fully reflect leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test. (Recommendation 12)

The Secretary of Defense should ensure that the Under Secretary of Defense for Acquisition and Sustainment revises its weapon system acquisition policies to require that such acquisition programs obtain and incorporate input from developmental and operational testers in the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads. (Recommendation 13)

Agency Comments and Our Evaluation

In June 2025, we provided a draft of this report to the Department of Defense (DOD) for review and comment. In November 2025, DOD provided written comments, reproduced in appendix II, in which it concurred with seven recommendations (recommendations 4-9 and 13), partially concurred with five recommendations (recommendations 1-3 and 10-11), and did not concur with one recommendation (recommendation 12). DOD also provided technical comments, which we incorporated, as appropriate.

In addition to the comments on our recommendations, DOD provided summary comments related to the overall findings of our report. These included statements of concurrence from the Army and the Air Force. They also included comments from the Office of the Under Secretary of Defense for Research and Engineering and the Navy that they deemed as "critical," with the Navy stating that it did not concur with the report as written.

We address these comments, as appropriate, below, including when discussing specific recommendations. Among their comments, DOD and the Navy stated that we had unfairly cited test and evaluation as causing acquisition delays. That was not our intent. We have clarified in the report that these delays stem from a combination of factors, including identifying technical and performance challenges discovered late in a program's development.

Additionally, DOD requested that we estimate the costs and savings associated with implementing our recommendations. We agree that DOD needs to weigh the costs associated with taking various courses of actions and needs to identify cost-effective solutions. The recommendations we identified, however, are intended to help DOD incorporate and maximize the use of practices used by leading commercial companies to deliver products and capabilities to their customers that meets their needs on a timely basis and at a price they can afford. These imperatives are echoed in the Secretary of Defense's November 2025 memorandum that called for transforming the defense acquisition system into the warfighting acquisition system to accelerate the fielding of urgently needed capabilities.[36]

[36]Secretary of Defense Memorandum, *Transforming the Defense Acquisition System into the Warfighting Acquisition System to Accelerate Fielding of Urgently Needed Capabilities to Our Warriors* (Nov. 7, 2025).

Below we organized our responses to DOD's comments on the recommendations by the individual agencies responsible for implementing them. The Army concurred with each of the three recommendations (recommendations 7-9) we made to fully reflect leading practices in its test and evaluation policy.

The Air Force concurred with each of the three recommendations (recommendations 4-6) we made for it to fully reflect leading practices in its test and evaluation policy. For recommendations 4 and 5 it cited existing policy, which we reviewed in our analysis, as sufficient to satisfy our recommendations. For example, in response to recommendation 4, the Air Force pointed to its existing policy as already supportive of our recommendation. Similarly, for recommendation 5, the Air Force pointed to its existing test and evaluation policy as fulfilling the part our recommendation that emphasizes tester input into test strategy and test plan development. As detailed earlier in the report, our analysis of Air Force's existing policy found that it did not fully reflect leading practices for product development. We therefore note, as the Air Force works to implement our recommendations, that incorporating the leading practices that this report identifies will likely necessitate substantive changes rather than marginal ones to existing policy.

DOD concurred with our recommendation directed to the Under Secretary of Defense for Acquisition and Sustainment (recommendation 13). This recommendation calls for the Under Secretary to revise DOD's weapon system acquisition policies to require weapon programs obtain and incorporate input from developmental and operational testers in the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads. Regarding recommendation 1 (discussed below), DOD indicated that it is constrained from implementing that recommendation until the Under Secretary has acted on recommendation 13. We believe this magnifies the importance of recommendation 13 and encourage the Under Secretary to move swiftly to implement the necessary policy revision.

DOD partially concurred with each of our three recommendations (recommendations 1-3) directed to the Under Secretary of Defense for Research and Engineering and the Director, Operational Test and Evaluation.

- For recommendation 1, DOD agreed that department-wide policies for developmental test and evaluation and operational test and evaluation require updates to better reflect leading practices. Among the potential updates, DOD stated that requiring test community input into the development of acquisition strategies would constitute a sound practice, but that such a requirement would exceed the current authority of DOD test and evaluation policies. DOD noted that it is constrained from implementing this recommendation until the DOD has addressed recommendation 13. As noted above, we are encouraged that the Under Secretary for Acquisition and Sustainment agreed to implement recommendation 13. We believe that once the Under Secretary has implemented recommendation 13 that this should allow DOD to update its test and evaluation policy, consistent with the revised acquisition policy and the intent of our recommendation.
- For recommendation 2, DOD agreed that iterative, integrated testing supports delivery of minimum viable products (MVP), but stated that MVPs are driven by acquisition strategies, not test plans or test strategies. Nonetheless, DOD committed to considering updates to test and evaluation policies to ensure the development of test strategies and test plans that include iterative and integrated testing approaches. DOD also said it would consider including language to leverage digital engineering tools across acquisition pathways. DOD's comments also reflect a misunderstanding of MVPs, including how leading companies define, develop, and test MVPs and the integral role they play in iterative development and testing. We are happy to reengage with DOD's test and evaluation leaders to further

discuss our work on key practices, including for test and evaluation, on which leading companies rely.[37] In our recent work on these leading practices, we found that the practices are essential to delivering relevant capabilities on timelines responsive to users' needs, consistent with DOD's own acquisition goals.

- For recommendation 3, DOD stated that the Under Secretary of Defense for Research and Engineering is working to update applicable policies and guidance to more similarly align with our recommendation. DOD stated that the Director concurs that it does not specify a process for how and when user feedback should be solicited through iterative testing, where applicable. DOD noted that any such process should be coordinated with the military services to ensure consistency across DOD-wide and service-level test and evaluation guidance. We encourage DOD to move out on these efforts and to document its resulting process updates in policy rather than in guidance to help ensure DOD achieves the consistency it seeks.
- Underpinning its partial concurrences across recommendations 1-3, DOD stated it did not believe incorporating testing-related leading practices within existing department-wide policies for systems engineering and digital engineering was warranted. We evaluated these two policies as part of our review at DOD's request. Specifically, in May 2025, DOD officials told us that the systems engineering and digital engineering policies contained many of the testing-related leading practices that we found lacking in the test and evaluation policies we reviewed. In our subsequent review of the systems engineering and digital engineering policies, we found the policies also did not reflect the leading practices related to test and evaluation. We agree with DOD's written comments that policies, including those for systems engineering, digital engineering, and test and evaluation, need not be duplicative. At the same time, given that we found leading practices missing from all the policies we reviewed, we offer that moving out on multiple fronts to drive leading practices into policy is wholly preferable to no movement at all. With this in mind, we continue to believe that DOD's systems engineering and digital engineering policies could be further improved to reflect the leading practices highlighted in this report.

The Navy partially agreed with two recommendations (recommendations 10-11) and did not agree with one recommendation (recommendation 12) we made to fully reflect leading practices in its test and evaluation policy.

- For recommendation 10, the Navy stated that it sees utility that digital threads, including digital twins and other modeling and simulation, bring to the requirements and acquisition processes and products. The Navy also stated that it is working to implement these as part of a digital strategy. However, the Navy indicated that it prefers to await updates to DOD acquisition policy on these topics before pursuing related changes to Navy acquisition policy, including test and evaluation policy. We point the Navy to the Secretary of Defense's November 2025 memorandum on accelerating delivery of capabilities to warfighters—a document that echoes several of the leading practices that we previously identified—and encourage it use that memorandum as a launch pad to initiate the changes that we have recommended.[38]
- For recommendation 11, the Navy pointed to its "Capabilities Based Test and Evaluation" process and its acquisition policy as supporting integrated testing and use of digital models, respectively. The Navy clarified that the digital models language in its policy does not specify digital threads or digital twins—

[37]See GAO-23-106222 and GAO-22-104513.

[38]Secretary of Defense Memorandum, *Transforming the Defense Acquisition System into the Warfighting Acquisition System to Accelerate Fielding of Urgently Needed Capabilities to Our Warriors* (Nov. 7, 2025).

and centers its planned actions to implement our recommendation around resolving that language difference. As with recommendation 10, discussed above, the Navy stated it plans to wait until DOD has updated its acquisition policies to specify digital threads or digital twins, before updating its test and evaluation policies. More importantly, the Navy's response did not address the crux of recommendation 11, which calls for policy changes to drive iterative testing approaches that support delivery of minimum viable products. As our prior work has found, digital threads and digital twins are key enablers for iterative testing and delivery of minimum viable products. At the same time, using these threads and twins to carry out traditional, linear-based test and evaluation activities is akin to using modern tools to carry out antiquated tasks—subsequently failing to capitalize on the full utility these tools can offer. We encourage the Navy, as it implements our recommendation, to break away from traditional approaches that have underperformed and incorporate iterative testing approaches underpinned by the use of digital twins and digital threads.

- For recommendation 12, the Navy did not concur and pointed to existing language in its test and evaluation policy that calls for weapon system acquisition programs to collaborate with resource sponsors, requirements officers, fleet operators, and operational testers to identify applicable mission tasks, measures, and conditions for the system under test. We agree with the Navy that it has this language in its policy; we reviewed it as part of our military department policy analysis. We disagree that this language connotes the equivalent of an end user agreement. End user agreements that detail a process for obtaining ongoing user input throughout testing ensure the capability under development is relevant and responsive to users' most critical needs. End user agreements are also fundamental to iterative testing approaches, which this report and our other prior reports have made clear are essential to delivering relevant capabilities on timelines responsive to users' needs.[39] By not requiring end user agreements under its test and evaluation policy, the Navy reinforces its existing linear testing approaches. These approaches constrain user feedback , which is crucial to the timely delivery of relevant capabilities to warfighters, as outlined in our prior work.[40]

We are sending copies of this report to the appropriate congressional committees. In addition, the report is available at no charge on the GAO website at https://www.gao.gov.

[39]See GAO-23-106222 and GAO-22-104513.

[40]See GAO-23-106222 and GAO-22-104513.

If you or your staff have any questions about this report, please contact me at oakleys@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made key contributions to the report are listed in appendix III.

//SIGNED//

Shelby S. Oakley
Director, Contracting and National Security Acquisitions

List of Committees

The Honorable Roger Wicker
Chairman
The Honorable Jack Reed
Ranking Member
Committee on Armed Services
United States Senate

The Honorable Mitch McConnell
Chair
The Honorable Christopher Coons
Ranking Member
Subcommittee on Defense
Committee on Appropriations
United States Senate

The Honorable Mike Rogers
Chairman
The Honorable Adam Smith
Ranking Member
Committee on Armed Services
House of Representatives

The Honorable Ken Calvert
Chairman
The Honorable Betty McCollum
Ranking Member
Subcommittee on Defense
Committee on Appropriations
House of Representatives

Appendix I: Objectives, Scope, and Methodology

This report (1) describes plans the Department of Defense (DOD) has established to modernize test and evaluation to deliver relevant capabilities faster to the warfighter, and (2) assesses the extent to which DOD-wide and military department policies for test and evaluation of weapon systems reflect our selected leading practices for product development.

For both objectives, we obtained documentation from DOD-wide and military department-level test and evaluation organizations and interviewed officials from those organizations to obtain additional context and corroborate our review and assessment of documentation. We conducted site visits to three of these test organizations, one at each of the military departments included in our review—Air Force, Army, and Navy. We selected these test organizations because they are primary test agencies at each of the military departments and their workforces were centrally located. At these site visits, we met with test officials and saw test and evaluation tools in practice, such as modeling and simulation and live virtual constructive environments. The DOD-wide and military department-level test and evaluation organizations we interviewed and visited include:

- Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E))
- Test Resource Management Center
- Office of the Director, Operational Test and Evaluation (DOT&E)
- Office of the Under Secretary of Defense for Acquisition and Sustainment
- Air Force Operational Test and Evaluation Center (AFOTEC)
- Air Force Test Center (AFTC) (site visit)
- Army Test and Evaluation Command (ATEC) (site visit)
- Navy Office of the Chief of Naval Operations
- Navy Operational Test and Evaluation Force (OPTEVFOR) (site visit)

To identify the plans DOD established to modernize test and evaluation to support rapid development of weapon systems, we interviewed officials in DOD-wide and military department-level test organizations and identified plans that they either developed or used to guide modernization of test and evaluation. We reviewed the DOD-wide and military department-level plans and identified both goals and concepts related to test and evaluation and common themes across the plans. The plans we reviewed include:

- Defense Science Board, *Test and Evaluation*, August 2024
- Defense Science Board, *Digital Engineering Capability to Automate Testing and Evaluation Final Report*, May 2024
- Office of the Secretary of Defense, *DOD Strategic Management Plan Fiscal Years 2022-2026*
- DOT&E, *Strategy Implementation Plan*, April 2023
- DOT&E, *Science and Technology Strategic Plan*, January 2021
- OUSD(R&E), *Test and Evaluation as a Continuum*, March 2023

- Office of the Deputy Assistant Secretary of Defense for Systems Engineering, *DOD Digital Engineering Strategy*, June 2018
- ATEC, *Fiscal Year 2024 ATEC Next: Campaign Plan for 2024-2029*
- ATEC, *Modernization Task Force Summary*, March 2024
- Under Secretary of the Army, *Army Digital Engineering Directive*, May 2024
- AFOTEC, *AFOTEC Strategic Vision 2030*, February 2024
- Air Force Materiel Command, *Air Force Materiel Command: An Accelerated Future State*
- AFTC, *AFTC Digital Modernization Strategy*, July 2022
- Office of the Deputy Assistant Secretary of the Navy for Research, Development, Test and Evaluation, *United States Navy & Marine Corps Digital Systems Engineering Transformation Strategy*, 2020
- OPTEVFOR, *OPTEVFOR Strategic Roadmap*, January 2022

We discussed test and evaluation modernization goals and the potential barriers to meeting those goals with officials at developmental and operational test organizations across DOD and the military departments, including the selected ones we visited.

To assess the extent to which DOD-wide and military department-level policies for test and evaluation reflect our selected leading practices for product development, we first identified relevant leading practices for product development and how these key practices apply to test and evaluation.[41] We then obtained and reviewed the DOD-wide test and evaluation policies and other related functional area policies on digital engineering and systems engineering. We also obtained and reviewed military department-level policies related to test and evaluation. We exclusively reviewed these test and evaluation policies, rather than guidance, because policies are governing documents that require courses of action for weapon system test and evaluation. Guidance, on the other hand, is not binding and is therefore optional to follow. The DOD-wide and military department-level policies include:

- Department of Defense Instruction 5000.89, *Test and Evaluation* (Nov. 19, 2020)
- Department of Defense Instruction 5000.98, *Operational Test and Evaluation and Live Fire Test and Evaluation* (Dec. 9, 2024)
- Department of Defense Instruction 5000.88, *Engineering of Defense Systems* (Nov. 18, 2020)
- Department of Defense Instruction 5000.97, *Digital Engineering* (Dec. 21, 2023)
- Department of the Air Force Instruction 99-103, *Capabilities-Based Test and Evaluation* (Dec. 9, 2021) (incorporating change, Feb. 28, 2025)
- Army Regulation 73-1, *Test and Evaluation Policy* (June 8, 2018)
- Secretary of the Navy Instruction 5000.2G, *Department of the Navy Implementation of the Defense Acquisition System and the Adaptative Acquisition Framework* (Apr. 8, 2022)

[41]GAO, *Leading Practices: Iterative Cycles Enable Rapid Delivery of Complex, Innovative Products*, GAO-23-106222 (Washington, D.C.: July 27, 2023).

We also reviewed draft test and evaluation policies from OUSD(R&E), the Air Force, and the Army. We assessed policies related to test and evaluation against the selected leading practices and relevant application to test and evaluation. We focused our review of the policies on language applicable to four pathways of the Adaptive Acquisition Framework: urgent capability acquisition, middle tier of acquisition, major capability acquisition, and software acquisition. We also shared preliminary observations with DOT&E, the Office of Developmental Test Evaluation & Assessment under OUSD(R&E), and military department officials.

In addition, we selected a nongeneralizable sample of eight ongoing weapon system acquisition programs as illustrative examples to examine the military departments' implementation of our selected leading practices for product development. To select the eight programs, we planned for a distribution of programs between the three primary military departments: Air Force, Army, and Navy. We identified major defense acquisition programs and middle tier of acquisition programs for which we had ongoing work that we could leverage. We excluded programs that were classified or programs for which significant amounts of programmatic information was sensitive. The team then narrowed down the selection based on whether the program was identified as potentially using modern design tools or test and evaluation approaches through: (1) interviews with test and evaluation officials at the OSD and military department-level, and (2) documentation from DOD and the Defense Science Board. The selected programs include:

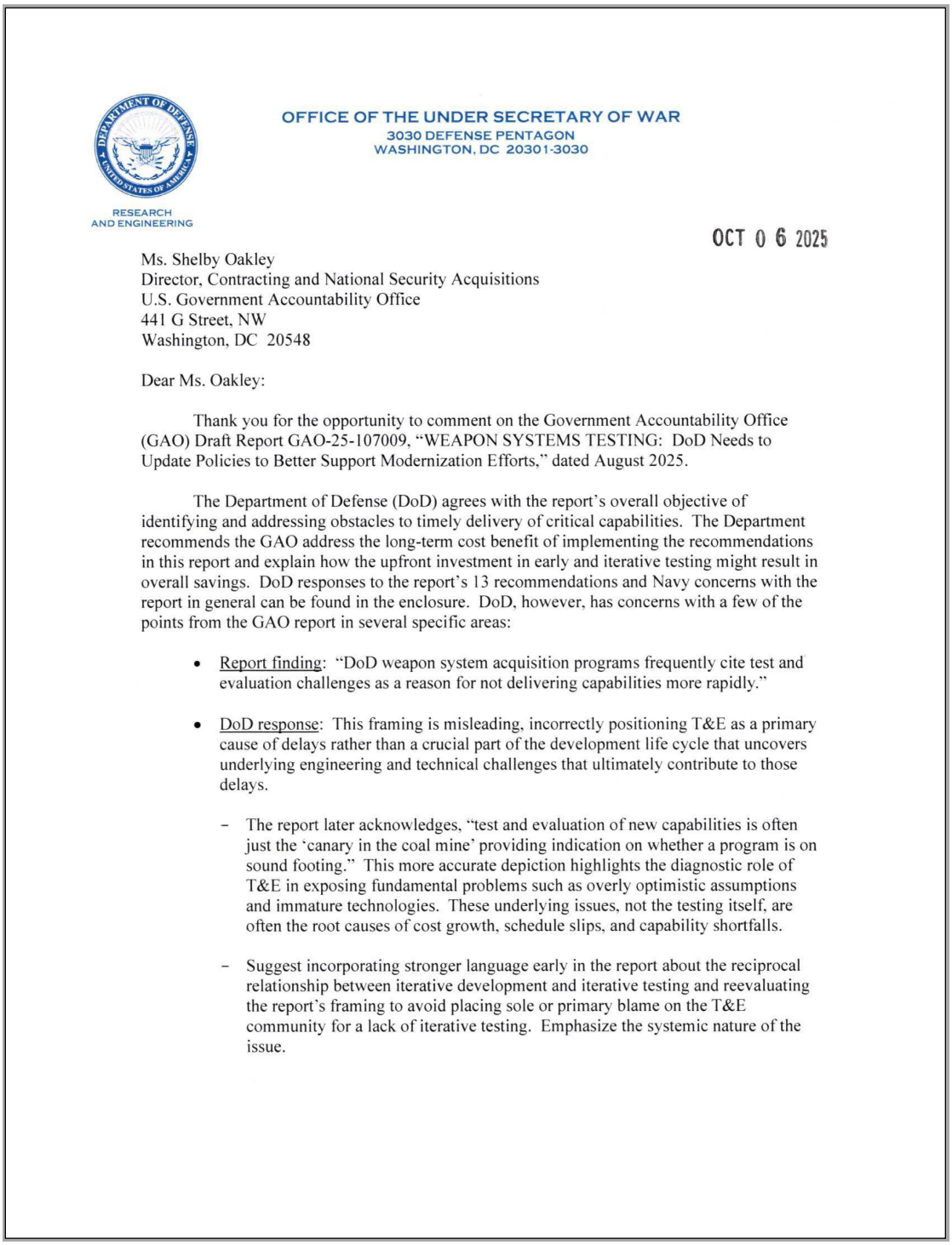
- Air Force:
  - T-7A Red Hawk trainer aircraft
  - F-22 Raptor fighter aircraft
  - F-35 Lightning II fighter aircraft
- Army:
  - Integrated Visual Augmentation System
  - XM30 Mechanized Infantry Combat Vehicle (formerly Optionally Manned Fighting Vehicle)
  - Future Long Range Assault Aircraft
- Navy:
  - CVN 78, *Gerald R. Ford* class nuclear aircraft carrier
  - Ship to Shore Connector amphibious craft

We obtained and reviewed program acquisition documents—including acquisition strategies and test plans—and other information. We reviewed the acquisition documents and analyzed whether they reflected elements of leading practices for product development as applied to the test and evaluation process.

We conducted this performance audit from August 2023 to December 2025 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Appendix II: Comments from the Department of Defense

DOD signed the letter in October 2025 and provided it to us in November 2025. Correspondingly, we changed our report number from GAO-25-107009 to GAO-26-107009.



OFFICE OF THE UNDER SECRETARY OF WAR
3030 DEFENSE PENTAGON
WASHINGTON, DC 20301-3030

RESEARCH
AND ENGINEERING

OCT 0 6 2025

Ms. Shelby Oakley
Director, Contracting and National Security Acquisitions
U.S. Government Accountability Office
441 G Street, NW
Washington, DC 20548

Dear Ms. Oakley:

Thank you for the opportunity to comment on the Government Accountability Office (GAO) Draft Report GAO-25-107009, "WEAPON SYSTEMS TESTING: DoD Needs to Update Policies to Better Support Modernization Efforts," dated August 2025.

The Department of Defense (DoD) agrees with the report's overall objective of identifying and addressing obstacles to timely delivery of critical capabilities. The Department recommends the GAO address the long-term cost benefit of implementing the recommendations in this report and explain how the upfront investment in early and iterative testing might result in overall savings. DoD responses to the report's 13 recommendations and Navy concerns with the report in general can be found in the enclosure. DoD, however, has concerns with a few of the points from the GAO report in several specific areas:

- Report finding: "DoD weapon system acquisition programs frequently cite test and evaluation challenges as a reason for not delivering capabilities more rapidly."

- DoD response: This framing is misleading, incorrectly positioning T&E as a primary cause of delays rather than a crucial part of the development life cycle that uncovers underlying engineering and technical challenges that ultimately contribute to those delays.

  - The report later acknowledges, "test and evaluation of new capabilities is often just the 'canary in the coal mine' providing indication on whether a program is on sound footing." This more accurate depiction highlights the diagnostic role of T&E in exposing fundamental problems such as overly optimistic assumptions and immature technologies. These underlying issues, not the testing itself, are often the root causes of cost growth, schedule slips, and capability shortfalls.

  - Suggest incorporating stronger language early in the report about the reciprocal relationship between iterative development and iterative testing and reevaluating the report's framing to avoid placing sole or primary blame on the T&E community for a lack of iterative testing. Emphasize the systemic nature of the issue.

- Report finding: "DoD and the military departments' plans to modernize test and evaluation share themes, including use of digital engineering tools and highly skilled workforces."

- DoD response: Concur, but these efforts should complement, not replace, a commitment to addressing the fundamental program issues that are identified through testing.

  - Digital engineering tools and processes offer valuable improvements in managing technical risk and promoting agility, but they will not eliminate the inherent challenges in developing the sophisticated cyber-physical systems necessary to maintain our military dominance. The level of effort required for such complex system development endeavors remains substantial.

  - The report should add a crucial qualifier about the independent validation of digital twins. A digital twin's utility for testing is directly dependent on its fidelity and validation.

- Report finding: "DoD's digital engineering policy and the test and evaluation section of DoD's system engineering policy do not describe specific processes to ensure application of leading practices to testing."

- DoD response: DoD digital engineering policy is comprehensive, outlining actions to enable faster, higher-quality decision making in weapon system design, development, testing, fielding, and sustainment. The DoD systems engineering policy provides guidance for a comprehensive engineering program for defense systems, encompassing the necessary engineering management activities. Specific application of digital engineering to T&E, however, belongs in the developmental and operational T&E policies.

- Report finding: "DoD has a unique opportunity to not only retool its existing test and evaluation enterprise, but to redefine the role that enterprise can play in enabling faster delivery of relevant capabilities to warfighters. Fully incorporating leading practices into policies relevant to weapon system test and evaluation could help pivot their current reactive role to a proactive one, informing and aiding defense acquisition efforts."

- DoD response: Concur on the need to adopt leading T&E practices. However, to truly drive changes in the implementation and execution of T&E, changes should come from updated acquisition policy before T&E policy can be updated to properly support an effective, efficient, and timely acquisition.

Specific Service responses to the draft report include the following:

- The Department of the Army concurs with the report with no comment.

- The Department of the Air Force concurs with the report with no comment.
- The Marine Corps consolidated its response into the Navy's response.
- The Department of the Navy (DON) T&E does not concur with the report as written. DON T&E discussed its concerns with the GAO report authors and was assured by those authors that the bulletized list below would be considered in the final version of the report.
  - CRITICAL COMMENT: T&E CAUSES PROGRAM DELAYS – DON T&E disagrees with GAO's framing of T&E challenges, under the GAO Highlights, as a reason for not delivering capabilities more rapidly. Specifically, the statement, "DoD weapon system acquisition programs frequently cite T&E challenges as a reason for not delivering capabilities more rapidly," appears to contradict the discussion of the value of T&E presented later in the report. Delays to system deliveries are primarily due to system issues uncovered during T&E, not as a result of the practice of T&E. Additionally, to our understanding, GAO did not quantify this assessment, yet the report introduction establishes this assumption as fact and as the problem that the report is trying to address. Request GAO provide factual references to support this conclusion or remove the statement.
  - CRITICAL COMMENT: INADEQUATE POLICY – Gaps in T&E policy, identified in the body of the report, do not accurately reflect existing DON acquisition policy. Secretary of the Navy Instruction (SECNAVINST) 5000.2G specifically addresses early user involvement, integrated developmental and operational testing, and the use of modeling and simulation (M&S) as part of the T&E process. DON T&E does agree that DoD T&E policies are handicapped by the segregation of operational T&E and live fire T&E policy from developmental T&E policy, which can cause confusion and misalignment and can preclude the coherent execution of DoD integrated testing policy required in today's environment.
  - CRITICAL COMMENT: CVN-78 FACTUAL ERRORS (page 35 regarding Figure 9, CVN 78 Iterative Testing) – Iteration within ship classes occurs over years and multiple classes of ships. Once ship construction has begun, capabilities are increased through hardware and software updates and tested in follow-on operational T&E.
    - Additionally, DON T&E is concerned with the statement, "However, we found that the Navy test plans we reviewed did not identify how user feedback can be used to support an iterative test process. As a result, programs across pathways do not ensure that the weapon system under test meets essential user needs." Specifically, the report is unclear on which test plans were deficient and what the impacts of those deficiencies were on the total T&E effort.
    - DON T&E also disagrees with the assessment that programs are not meeting essential user needs. In the execution of DON T&E, users and user

representatives are part of the Mission Based Test Design process up front for both the Navy and Marine Corps. Their inclusion in this process directly supports operational T&E to ensure user needs are addressed in support of mission execution.

- CRITICAL COMMENT: REQUIREMENT FOR T&E TO DRIVE PROGRAM OFFICE IMPLEMENTATION OF DIGITAL TWINS AND THREADS INTO ACQUISITION STRATEGIES AND T&E PROCESSES – DON T&E disagrees with the requirement for T&E to drive program office implementation of digital twins and threads into acquisition strategies and T&E processes.

  - DON T&E assesses that a program should not make an investment in digital twins and digital threads solely for the purpose of T&E. M&S and data solutions should support the entire program from requirement development, to engineering, potentially to tactics development and training, and through sustainment. The program T&E team can make recommendations to the program manager on the potential value of M&S tools, including digital twins, as well as data storage solutions in supporting an efficient and effective T&E program. However, the program manager makes the decisions to invest in resources and determines contract deliverables, including M&S tools and data solutions. Once those decisions are made, the T&E team will use the resources provided to support applicable T&E execution and data collection.

  - Recommend this report align with the findings of the Defense Business Board (DBB) report DBB Fiscal Year 2024-2003, "Creating a Digital Ecosystem." This report shows that DoD's data policies should be driven from the top down rather than up from T&E acquisition policy.

My point of contact is Mr. Orlando Flores, Acting Director, Developmental Test, Evaluation, and Assessments, at orlando.f.flores.civ@mail.mil or 571-372-4145.

Sincerely,

Joseph P. Morici
Performing the Duties of the
Deputy Under Secretary of War
for Research and Engineering

Enclosure:
As stated

4

GAO Draft Report Dated August 2025
GAO-25-107009 (GAO Code 107009)

"WEAPON SYSTEMS TESTING: DoD Needs to Update Policies to Better Support Modernization Efforts"

Department of Defense Comments to the GAO Recommendations

Recommendation 1: The Secretary of War should ensure that the Under Secretary of War for Research and Engineering and the Director of Operational Test and Evaluation revise their weapon system test and evaluation, digital engineering, and systems engineering policies to fully reflect leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads.

DoD Response: The Department of Defense (DoD) partially concurs with this recommendation. DoD agrees that test and evaluation (T&E) policies for developmental test and evaluation (DoD Instruction (DoDI) 5000.89) and operational test and evaluation (DoDI 5000.98) should be updated to better reflect leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies. Although requiring test community input into the development of acquisition strategies would be a sound practice, the recommendation exceeds the current authority of DoD T&E policies. The report should reflect these limitations, particularly for developmental test. DoD policy changes directing the development of acquisition strategies are under the purview of the Office of the Under Secretary of War for Acquisition and Sustainment (OUSW(A&S)) as noted in recommendation #13. The ability to meet the intent of this recommendation is dependent on OUSW(A&S) execution of recommendation #13 and, to ensure consistency, recommendation #13 should be linked to the report's iterative development themes. The report should explicitly address the interplay and potential conflicts between T&E policy and acquisition policy regarding mandatory T&E involvement in early acquisition decisions. Consider adding a section or appendix that clearly delineates the policy landscape and identifies where T&E policy recommendations might require corresponding changes in acquisition policy to be fully effective. Also, rather than suggesting that policy should allow testers to "advocate for access to digital twins," consider strengthening the language to indicate that policy should "require/suggest access to digital twins for test and evaluation use cases" to ensure integration.

Additionally, DoDI 5000.98 mentions the use of digital twins but not digital threads in operational testing. A key prerequisite of establishing access to technologies such as digital twins and digital threads is that OUSW(A&S) policy should direct effective contracting to enable the test community to obtain technical data packages (for platforms) and source code (for software-centric solutions and applications).

However, DoD does not agree with the recommendation to revise the policies for digital engineering (DoDI 5000.97, "Digital Engineering") and systems engineering (DoDI 5000.88,

"Engineering of Defense Systems") to address this level of detail. DoDI 5000.88 focuses on systems engineering principles and life cycle integration, and DoDI 5000.97 provides overarching guidance on digital engineering practices. Incorporating specific requirements for tester involvement in acquisition strategies is outside the scope of these policies. These issuances appropriately reference T&E policies, ensuring alignment without duplicating content.

Furthermore, the representation of DoDI 5000.88 and DoDI 5000.97 in Figure 4 of the report, which focuses on explicit T&E practices, could be improved. Rather than depicting these policies as simply not addressing the leading practices, Figure 4 should acknowledge their role in enabling such practices through broader systems and digital engineering principles. DoDI 5000.88 and DoDI 5000.97 establish the foundational framework and requirements for systems engineering and digital engineering, which directly enable and support the integration of robust and iterative testing throughout the acquisition life cycle, as detailed below.

- DoDI 5000.88, "Engineering of Defense Systems," requires a digital engineering implementation plan within the Systems Engineering Plan, ensuring the development of a continuous end-to-end digital representation of the system and the establishment of a digital authoritative source of truth. This policy also requires the program manager to make the relevant digital models available throughout the program life cycle, thereby supporting robust testing and informed decision making by all stakeholders, including testers.

- DoDI 5000.97, "Digital Engineering," is a comprehensive policy that outlines the actions needed to enable faster, higher-quality decision making in weapon system design, development, testing, fielding, and sustainment. It mandates addressing digital engineering in the acquisition strategy, including its life cycle implementation and expected benefits. This policy further reinforces the need for technically accurate digital models at appropriate fidelity levels to ensure realistic test environments, promoting the very integration that GAO recommends. This policy, as highlighted in GAO-25-107263, encourages the use of digital engineering tools, including digital twins and digital threads.

Recommendation 2: The Secretary of War should ensure that the Under Secretary of War for Research and Engineering and the Director of Operational Test and Evaluation revise their weapon system test and evaluation, digital engineering, and systems engineering policies to fully reflect leading practices for product development by requiring developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products.

DoD Response: The Department of Defense (DoD) partially concurs with this recommendation. DoD agrees that the policies for developmental test and evaluation (DoD Instruction (DoDI) 5000.89) and operational test and evaluation (DoDI 5000.98) could be updated to better address the recommendation. DoD agrees that iterative, integrated testing in current policy supports delivery of minimum viable products (MVPs), but the MVP is driven by acquisition strategies and not T&E plans. Additionally, apart from software, operational testing does not result in MVPs. The Under Secretary of War for Research and Engineering (USW(R&E)) and the Director of Operational Test and Evaluation will consider updates to these test and evaluation (T&E) policies to ensure the development of test strategies and test plans that include iterative

and integrated testing approaches leveraging digital engineering tools across acquisition pathways. Upgrading from MVP to full capability is much easier said than done in both DoD and the private sector—and it gets very hard very quickly once the human factor comes into play. Perhaps the report can address practical implementation challenges (hardware/legacy systems) and provide more nuanced discussions on the practical challenges of implementing iterative development and testing, particularly for complex hardware systems and legacy DoD software systems. Recommend that the report include case studies or examples that illustrate successful iterative development and testing in challenging DoD environments or offer strategies for mitigating the difficulties identified in the comments.

Also, DoD does not agree with the recommendation to update the policies for digital engineering (DoDI 5000.97) and systems engineering (DoDI 5000.88) to address this level of detail. Incorporating specific requirements for developing test strategies and test plans is outside the scope of these policies. These issuances appropriately reference the T&E policies, ensuring alignment without duplicating content. Although the Department disagrees with incorporating this level of detail into the digital engineering and systems engineering policies, the USW(R&E) will consider broader applications of MVP and agile system development approaches across DoD systems engineering, beyond software engineering and across acquisition pathways, as appropriate in future updates to DoDI 5000.88 or other issuances. Suggest clarifying the Figure 4 legend to align with the figure's content. Consider updating the legend to explicitly state "Application of leading practice to test and evaluation" or adjusting the shading of those policies to "partially reflects" to avoid potential misinterpretation.

Recommendation 3: The Secretary of War should ensure that the USW(R&E) and the Director of Operational Test and Evaluation revise their weapon system test and evaluation, digital engineering, and systems engineering policies to fully reflect leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test.

DoD Response: The DoD partially concurs with this recommendation. The USW(R&E) is working to update applicable policies and guidance similar to this recommendation, promoting early engagement to support and be informed by a mission engineering process with mission threads sourced from end users, moving test and evaluation (T&E) from a serial set of activities to a new integrative framework focused on a continuum of activities driving continual decision support across the capability life cycle.

In addition, operational testing requires input and feedback from users throughout the test planning process; Director of Operational Test and Evaluation (DOT&E) policies require the use of soldiers, sailors, airmen, and marines in operational tests for operational realism. DoD Instruction (DoDI) 5000.98 states that system users are a part of the T&E working integrated product teams; DOT&E concurs that it does not specify a process for how and when user feedback should be solicited through iterative testing, where applicable. Such a process should be coordinated with the Services so both DoD-wide and Service-level T&E guidance is updated to reflect the same process.

However, DoD does not agree with the recommendation to update the policies for digital engineering (DoDI 5000.97) and systems engineering (DoDI 5000.88) to address this level of detail. Incorporating specific requirements for developing test strategies and test plans is outside the scope of these policies. These issuances appropriately reference the T&E policies, ensuring alignment without duplicating content. Furthermore, DoDI 5000.88 currently requires lead systems engineers to conduct frequent and iterative end user validation of features and usability for identifying, communicating, and visualizing user needs under defined operational conditions and expected mission threads.

Recommendation 4: The Secretary of the Air Force should, as the Air Force revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads.

DoD Response: The Air Force concurs with the intent of this recommendation. The Air Force concurs that early tester involvement in acquisition planning is an essential leading practice, and current Air Force policy (DoD Instruction 5000.89/Department of the Air Force Instruction (DAFI) 99-103, December 9, 2021) is written to support this thinking. In addition, current Air Force publications refer to the use of digital twins and digital threads in test planning and evaluation of systems. The current DAFI 99-103 references digital twins and digital threads in its current release. The Air Force will continue to strengthen policy to ensure the Service achieves maximum benefit from previous and future efforts.

Recommendation 5: The Secretary of the Air Force should, as the Air Force revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products.

DoD Response: The Air Force concurs with comment on this recommendation. DoD Instruction (DoDI) 5000.89/Department of the Air Force Instruction (DAFI) 99-103 already requires "developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach." The decision to invest in modeling and simulation such as digital twins resides with the program manager. Delivery of minimum viable products outside of agile software development is a new concept, but minimum viable product and associated capabilities are determined by the user and program manager.

Recommendation 6: The Secretary of the Air Force should, as the Air Force revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test.

DoD Response: The Air Force concurs with comment on this recommendation. The Air Force obtains user feedback during both developmental and operational test and evaluation by having users execute test for and with Operational Test Agencies: Air Force Operational Test and

Evaluation Center (AFOTEC) and Space Training and Readiness Command (STARCOM). Users also provide recommended changes to systems with AF Form 1067, "Modification Proposal." The Air Force will look to incorporate the engagement of end users during developmental test in future policy updates.

Recommendation 7: The Secretary of the Army should, as the Army revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads.

DoD Response: The Army concurs with this recommendation.

Recommendation 8: The Secretary of the Army should, as the Army revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products.

DoD Response: The Army concurs with this recommendation.

Recommendation 9: The Secretary of the Army should, as the Army revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test.

DoD Response: The Army concurs with this recommendation.

Recommendation 10: The Secretary of the Navy should revise the military department's weapon system test and evaluation policy to fully reflect leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads.

DoD Response: PARTIALLY CONCUR. Department of the Navy (DON) Test and Evaluation (T&E) concurs with the utility that digital threads, including digital twins and other modeling and simulation, bring to the requirements and acquisition processes and products and is working to implement these as part of its digital strategy. Once the Department of Defense issues specific language to describe these activities (including generating acquisition strategies) in its acquisition policy documents, DON will incorporate those changes in the next update to Secretary of the Navy Instruction (SECNAVINST) 5000.2G acquisition policy, which includes T&E policy.

Recommendation 11: The Secretary of the Navy should revise the military department's weapon system test and evaluation policy to fully reflect leading practices for product development by requiring developmental and operational testers to develop test strategies and

test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products.

DoD Response: PARTIALLY CONCUR – Existing Department of the Navy (DON) Test and Evaluation policy, included in the DON's acquisition instruction, Secretary of the Navy Instruction (SECNAVINST) 5000.2G, currently directs programs to conduct Capabilities Based Test and Evaluation (CBTE), which is based on the warfighter tasks and conditions identified in the Mission Based Test Design, and maps all testing and identified deficiencies to mission capabilities and mission impact. CBTE institutionalizes integrated testing. SECNAVINST 5000.2G also provides a requirement for building digital models and sharing data but does not specify digital threads or digital twins. Once the Department of Defense issues specific language to describe these activities in its respective acquisition policy documents, DON will incorporate those changes in the next update to SECNAVINST 5000.2G.

Recommendation 12: The Secretary of the Navy should revise the military department's weapon system test and evaluation policy to fully reflect leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test.

DoD Response: DO NOT CONCUR – Existing Department of the Navy acquisition policy states, "the program collaborates with Resource Sponsor/Requirements Office/Fleet/Force/OTA to identify applicable mission tasks, measures and conditions, as derived from applicable documents (including Required Operational Capabilities (ROC) and Projected Operational Environment (POE), Naval Tactical Tasks, Marine Corps Tasks, Coast Guard Tasks, etc.) for the System Under Test." The report is unclear in establishing what additional end user coordination would accelerate delivery of warfighting capability.

Recommendation 13: The Secretary of War should ensure that the Under Secretary of War for Acquisition and Sustainment revise its weapon system acquisition policies to require that such acquisition programs obtain and incorporate input from developmental and operational testers in the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads.

DoD Response: The Office of the Under Secretary of War for Acquisition and Sustainment concurs with this recommendation.

Accessible Text for Appendix II: Comments from the Department of Defense

OFFICE OF THE UNDER SECRETARY OF WAR
3030 DEFENSE PENTAGON
WASHINGTON, DC 20301-3030

RESEARCH AND ENGINEERING

OCT 06 2025

Ms. Shelby Oakley
Director, Contracting and National Security Acquisitions
U.S. Government Accountability Office
441 G Street, NW
Washington, DC 20548

Dear Ms. Oakley:

Thank you for the opportunity to comment on the Government Accountability Office (GAO) Draft Report GAO-25-107009, "WEAPON SYSTEMS TESTING: DoD Needs to Update Policies to Better Support Modernization Efforts," dated August 2025.

The Department of Defense (DoD) agrees with the report's overall objective of identifying and addressing obstacles to timely delivery of critical capabilities. The Department recommends the GAO address the long-term cost benefit of implementing the recommendations in this report and explain how the upfront investment in early and iterative testing might result in overall savings. DoD responses to the report's 13 recommendations and Navy concerns with the report in general can be found in the enclosure. DoD. however, has concerns with a few of the points from the GAO report in several specific areas:

- Report finding: "DoD weapon system acquisition programs frequently cite test and evaluation challenges as a reason for not delivering capabilities more rapidly."

- DoD response: This framing is misleading, incorrectly positioning T&E as a primary cause of delays rather than a crucial part of the development life cycle that uncovers underlying engineering and technical challenges that ultimately contribute to those delays.

    - The report later acknowledges, "test and evaluation of new capabilities is often just the ·canary in the coal mine' providing indication on whether a program is on sound footing." This more accurate depiction highlights the diagnostic role of T&E in exposing fundamental problems such as overly optimistic assumptions and immature technologies. These underlying issues, not the testing itself, are often the root causes of cost growth, schedule slips, and capability shortfalls.

    - Suggest incorporating stronger language early in the report about the reciprocal relationship between iterative development and iterative testing and reevaluating the report's framing to avoid placing sole or primary blame on the T&E community for a lack of iterative testing. Emphasize the systemic nature of the issue.

- Report finding: "DoD and the military departments' plans to modernize test and evaluation share themes, including use of digital engineering tools and highly skilled workforces."

- DoD response: Concur, but these efforts should complement, not replace, a commitment to addressing the fundamental program issues that are identified through testing.

    - Digital engineering tools and processes offer valuable improvements in managing technical risk and promoting agility, but they will not eliminate the inherent challenges in developing the sophisticated cyber-physical systems necessary to maintain our military dominance. The level of effort required for such complex system development endeavors remains substantial.

    - The report should add a crucial qualifier about the independent validation of digital twins. A digital twin's utility for testing is directly dependent on its fidelity and validation.

- Report finding: "DoD's digital engineering policy and the test and evaluation section of DoD's system engineering policy do not describe specific processes to ensure application of leading practices to testing."

- DoD response: DoD digital engineering policy is comprehensive, outlining actions to enable faster, higher-quality decision making in weapon system design, development, testing, fielding, and sustainment. The DoD systems engineering policy provides guidance for a comprehensive engineering program for defense systems, encompassing the necessary engineering management activities. Specific application of digital engineering to T&E, however, belongs in the developmental and operational T&E policies.

- Report finding: "DoD has a unique opportunity to not only retool its existing test and evaluation enterprise, but to redefine the role that enterprise can play in enabling faster delivery of relevant capabilities to warfighters. Fully incorporating leading practices into policies relevant to weapon system test and evaluation could help pivot their current reactive role to a proactive one, informing and aiding defense acquisition efforts."

- DoD response: Concur on the need to adopt leading T&E practices. However, to truly drive changes in the implementation and execution of T&E, changes should come from updated acquisition policy before T&E policy can be updated to properly support an effective, efficient, and timely acquisition.

Specific Service responses to the draft report include the following:

- The Department of the Army concurs with the report with no comment.

- The Department of the Air Force concurs with the report with no comment.

- The Marine Corps consolidated its response into the Navy's response.
- The Department of the Navy (DON) T&E does not concur with the report as written. DON T&E discussed its concerns with the GAO report authors and was assured by those authors that the bulletized list below would be considered in the final version of the report.
    - CRITICAL COMMENT: T&E CAUSES PROGRAM DELAYS - DON T&E disagrees with GAO's framing of T&E challenges, under the GAO Highlights, as a reason for not delivering capabilities more rapidly. Specifically, the statement, "DoD weapon system acquisition programs frequently cite T&E challenges as a reason for not delivering capabilities more rapidly," appears to contradict the discussion of the value of T&E presented later in the report. Delays to system deliveries are primarily due to system issues uncovered during T&E, not as a result of the practice of T&E. Additionally, to our understanding, GAO did not quantify this assessment, yet the report introduction establishes this assumption as fact and as the problem that the report is trying to address. Request GAO provide factual references to support this conclusion or remove the statement.
    - CRITICAL COMMENT: INADEQUATE POLICY -Gaps in T&E policy, identified in the body of the report, do not accurately reflect existing DON acquisition policy. Secretary of the Navy Instruction (SECNAVINST) 5000.2G specifically addresses early user involvement, integrated developmental and operational testing, and the use of modeling and simulation (M&S) as part of the T&E process. DON T&E does agree that DoD T&E policies are handicapped by the segregation of operational T&E and live fire T&E policy from developmental T&E policy, which can cause confusion and misalignment and can preclude the coherent execution of DoD integrated testing policy required in today's environment.
    - CRITICAL COMMENT: CVN-78 FACTUAL ERRORS (page 35 regarding Figure 9, CVN 78 Iterative Testing) - Iteration within ship classes occurs over years and multiple classes of ships. Once ship construction has begun, capabilities are increased through hardware and software updates and tested in follow-on operational T&E.
        - Additionally, DON T&E is concerned with the statement, "However, we found that the Navy test plans we reviewed did not identify how user feedback can be used to support an iterative test process. As a result, programs across pathways do not ensure that the weapon system under test meets essential user needs." Specifically, the report is unclear on which test plans were deficient and what the impacts of those deficiencies were on the total T&E effort.
        - DON T&E also disagrees with the assessment that programs are not meeting essential user needs. In the execution of DON T&E, users and user representatives are part of the Mission Based Test Design process up front for both the Navy and Marine Corps. Their inclusion in this process directly supports operational T&E to ensure user needs are addressed in support of mission execution.
    - CRITICAL COMMENT: REQUIREMENT FOR T&E TO DRIVE PROGRAM OFFICE IMPLEMENTATION OF DIGITAL TWINS AND THREADS INTO ACQUISITION STRATEGIES

AND T&E PROCESSES - DON T&E disagrees with the requirement for T&E to drive program office implementation of digital twins and threads into acquisition strategies and T&E processes.

- DON T&E assesses that a program should not make an investment in digital twins and digital threads solely for the purpose of T&E. M&S and data solutions should support the entire program from requirement development, to engineering, potentially to tactics development and training, and through sustainment. The program T&E team can make recommendations to the program manager on the potential value of M&S tools, including digital twins, as well as data storage solutions in supporting an efficient and effective T&E program. However, the program manager makes the decisions to invest in resources and determines contract deliverables, including M&S tools and data solutions. Once those decisions are made, the T&E team will use the resources provided to support applicable T&E execution and data collection.

- Recommend this report align with the findings of the Defense Business Board (DBB) report DBB Fiscal Year 2024-2003, "Creating a Digital Ecosystem." This report shows that DoD's data policies should be driven from the top down rather than up from T&E acquisition policy.

My point of contact is Mr. Orlando Flores, Acting Director, Developmental Test, Evaluation, and Assessments, at orlando.f.flores.civ@mail.mil or 571-372-4145.

Sincerely,

Joseph P. Morici
Performing the Duties of the
Deputy Under Secretary of War
for Research and Engineering

Enclosure:
As stated

GAO Draft Report Dated August 2025 GAO-25-107009 (GAO Code 107009)

"WEAPON SYSTEMS TESTING: DoD Needs to Update Policies to Better Support Modernization Efforts"

Department of Defense Comments to the GAO Recommendations

Recommendation 1: The Secretary of War should ensure that the Under Secretary of War for Research and Engineering and the Director of Operational Test and Evaluation revise their weapon system test and evaluation, digital engineering, and systems engineering policies to fully reflect leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads.

DoD Response: The Department of Defense (DoD) partially concurs with this recommendation. DoD agrees that test and evaluation (T&E) policies for developmental test and evaluation (DoD Instruction (DoDI) 5000.89) and operational test and evaluation (DoDI 5000.98) should be updated to better reflect leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies. Although requiring test community input into the development of acquisition strategies would be a sound practice, the recommendation exceeds the current authority of DoD T&E policies. The report should reflect these limitations, particularly for developmental test. DoD policy changes directing the development of acquisition strategies are under the purview of the Office of the Under Secretary of War for Acquisition and Sustainment (OUSW(A&S)) as noted in recommendation #13. The ability to meet the intent of this recommendation is dependent on OUSW(A&S) execution of recommendation #13 and, to ensure consistency, recommendation #13 should be linked to the report's iterative development themes. The report should explicitly address the interplay and potential conflicts between T&E policy and acquisition policy regarding mandatory T&E involvement in early acquisition decisions. Consider adding a section or appendix that clearly delineates the policy landscape and identifies where T&E policy recommendations might require corresponding changes in acquisition policy to be fully effective. Also, rather than suggesting that policy should allow testers to "advocate for access to digital twins," consider strengthening the language to indicate that policy should "require/suggest access to digital twins for test and evaluation use cases" to ensure integration.

Additionally, DoDI 5000.98 mentions the use of digital twins but not digital threads in operational testing. A key prerequisite of establishing access to technologies such as digital twins and digital threads is that OUSW(A&S) policy should direct effective contracting to enable the test community to obtain technical data packages (for platforms) and source code (for software-centric solutions and applications).

However, DoD does not agree with the recommendation to revise the policies for digital engineering (DoDI 5000.97, "Digital Engineering") and systems engineering (DoDI 5000.88, "Engineering of Defense Systems") to address this level of detail. DoDI 5000.88 focuses on systems engineering principles and life cycle integration, and DoDI 5000.97 provides overarching guidance on digital engineering practices. Incorporating specific requirements for tester involvement in acquisition strategies is outside the scope of these policies. These issuances appropriately reference T&E policies, ensuring alignment without duplicating content.

Furthermore, the representation of DoDI 5000.88 and DoDI 5000.97 in Figure 4 of the report, which focuses on explicit T&E practices, could be improved. Rather than depicting these policies as simply not addressing the leading practices, Figure 4 should acknowledge their role in enabling such practices through broader systems and digital engineering principles. DoDI 5000.88 and DoDI 5000.97 establish the foundational framework and requirements for systems engineering and digital engineering, which directly enable and support the integration of robust and iterative testing throughout the acquisition life cycle, as detailed below.

- DoDI 5000.88, "Engineering of Defense Systems," requires a digital engineering implementation plan within the Systems Engineering Plan, ensuring the development of a continuous end-to-end digital representation of the system and the establishment of a digital authoritative source of truth. This policy also requires the program manager to make the relevant digital models available throughout the program life cycle, thereby supporting robust testing and informed decision making by all stakeholders, including testers.
- DoDI 5000.97, "Digital Engineering," is a comprehensive policy that outlines the actions needed to enable faster, higher-quality decision making in weapon system design, development, testing, fielding, and sustainment. It mandates addressing digital engineering in the acquisition strategy, including its life cycle implementation and expected benefits. This policy further reinforces the need for technically accurate

digital models at appropriate fidelity levels to ensure realistic test environments, promoting the very integration that GAO recommends. This policy, as highlighted in GAO-25-107263, encourages the use of digital engineering tools, including digital twins and digital threads.

Recommendation 2: The Secretary of War should ensure that the Under Secretary of War for Research and Engineering and the Director of Operational Test and Evaluation revise their weapon system test and evaluation, digital engineering, and systems engineering policies to fully reflect leading practices for product development by requiring developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products.

DoD Response: The Department of Defense (DoD) partially concurs with this recommendation. DoD agrees that the policies for developmental test and evaluation (DoD Instruction (DoDI) 5000.89) and operational test and evaluation (DoDI 5000.98) could be updated to better address the recommendation. DoD agrees that iterative, integrated testing in current policy supports delivery of minimum viable products (MVPs), but the MVP is driven by acquisition strategies and not T&E plans. Additionally, apart from software, operational testing does not result in MVPs. The Under Secretary of War for Research and Engineering (USW(R&E)) and the Director of Operational Test and Evaluation will consider updates to these test and evaluation (T&E) policies to ensure the development of test strategies and test plans that include iterative and integrated testing approaches leveraging digital engineering tools across acquisition pathways. Upgrading from MVP to full capability is much easier said than done in both DoD and the private sector—and it gets very hard very quickly once the human factor comes into play. Perhaps the report can address practical implementation challenges (hardware/legacy systems) and provide more nuanced discussions on the practical challenges of implementing iterative development and testing, particularly for complex hardware systems and legacy DoD software systems. Recommend that the report include case studies or examples that illustrate successful iterative development and testing in challenging DoD environments or offer strategies for mitigating the difficulties identified in the comments.

Also, DoD does not agree with the recommendation to update the policies for digital engineering (DoDI 5000.97) and systems engineering (DoDI 5000.88) to address this level of detail. Incorporating specific requirements for developing test strategies and test plans is outside the scope of these policies. These issuances appropriately reference the T&E policies, ensuring alignment without duplicating content. Although the Department disagrees with incorporating this level of detail into the digital engineering and systems engineering policies, the USW(R&E) will consider broader applications of MVP and agile system development approaches across DoD systems engineering, beyond software engineering and across acquisition pathways, as appropriate in future updates to DoDI 5000.88 or other issuances. Suggest clarifying the Figure 4 legend to align with the figure's content. Consider updating the legend to explicitly state "Application of leading practice to test and evaluation" or adjusting the shading of those policies to "partially reflects" to avoid potential misinterpretation.

Recommendation 3: The Secretary of War should ensure that the USW(R&E) and the Director of Operational Test and Evaluation revise their weapon system test and evaluation, digital engineering, and systems engineering policies to fully reflect leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test.

DoD Response: The DoD partially concurs with this recommendation. The USW(R&E) is working to update applicable policies and guidance similar to this recommendation, promoting early engagement to support and be informed by a mission engineering process with mission threads sourced from end users, moving test and evaluation (T&E) from a serial set of activities to a new integrative framework focused on a continuum of activities driving continual decision support across the capability life cycle.

In addition, operational testing requires input and feedback from users throughout the test planning process; Director of Operational Test and Evaluation (DOT&E) policies require the use of soldiers, sailors, airmen, and marines in operational tests for operational realism. DoD Instruction (DoDI) 5000.98 states that system users are a part of the T&E working integrated product teams; DOT&E concurs that it does not specify a process for how and when user feedback should be solicited through iterative testing, where applicable. Such a process should be coordinated with the Services so both DoD-wide and Service-level T&E guidance is updated to reflect the same process.

However, DoD does not agree with the recommendation to update the policies for digital engineering (DoDI 5000.97) and systems engineering (DoDI 5000.88) to address this level of detail. Incorporating specific requirements for developing test strategies and test plans is outside the scope of these policies. These issuances appropriately reference the T&E policies, ensuring alignment without duplicating content. Furthermore, DoDI 5000.88 currently requires lead systems engineers to conduct frequent and iterative end user validation of features and usability for identifying, communicating, and visualizing user needs under defined operational conditions and expected mission threads.

Recommendation 4: The Secretary of the Air Force should, as the Air Force revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads.

DoD Response: The Air Force concurs with the intent of this recommendation. The Air Force concurs that early tester involvement in acquisition planning is an essential leading practice, and current Air Force policy (DoD Instruction 5000.89/Department of the Air Force Instruction (DAFI) 99-103, December 9, 2021) is written to support this thinking. In addition, current Air Force publications refer to the use of digital twins and digital threads in test planning and evaluation of systems. The current DAFI 99-103 references digital twins and digital threads in its current release. The Air Force will continue to strengthen policy to ensure the Service achieves maximum benefit from previous and future efforts.

Recommendation 5: The Secretary of the Air Force should, as the Air Force revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products.

DoD Response: The Air Force concurs with comment on this recommendation. DoD Instruction (DoDI) 5000.89/Department of the Air Force Instruction (DAFI) 99-103 already requires “developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach.” The decision to invest in modeling and simulation such as digital twins resides with the program

manager. Delivery of minimum viable products outside of agile software development is a new concept, but minimum viable product and associated capabilities are determined by the user and program manager.

Recommendation 6: The Secretary of the Air Force should, as the Air Force revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test.

DoD Response: The Air Force concurs with comment on this recommendation. The Air Force obtains user feedback during both developmental and operational test and evaluation by having users execute test for and with Operational Test Agencies: Air Force Operational Test and Evaluation Center (AFOTEC) and Space Training and Readiness Command (STARCOM). Users also provide recommended changes to systems with AF Form 1067, “Modification Proposal.” The Air Force will look to incorporate the engagement of end users during developmental test in future policy updates.

Recommendation 7: The Secretary of the Army should, as the Army revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads.

DoD Response: The Army concurs with this recommendation.

Recommendation 8: The Secretary of the Army should, as the Army revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products.

DoD Response: The Army concurs with this recommendation.

Recommendation 9: The Secretary of the Army should, as the Army revises its weapon system test and evaluation policy, ensure that policy fully reflects leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test.

DoD Response: The Army concurs with this recommendation.

Recommendation 10: The Secretary of the Navy should revise the military department’s weapon system test and evaluation policy to fully reflect leading practices for product development by requiring developmental and operational testers to provide input into the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads.

DoD Response: PARTIALLY CONCUR. Department of the Navy (DON) Test and Evaluation (T&E) concurs with the utility that digital threads, including digital twins and other modeling and simulation, bring to the

requirements and acquisition processes and products and is working to implement these as part of its digital strategy. Once the Department of Defense issues specific language to describe these activities (including generating acquisition strategies) in its acquisition policy documents, DON will incorporate those changes in the next update to Secretary of the Navy Instruction (SECNAVINST) 5000.2G acquisition policy, which includes T&E policy.

Recommendation 11: The Secretary of the Navy should revise the military department's weapon system test and evaluation policy to fully reflect leading practices for product development by requiring developmental and operational testers to develop test strategies and test plans that reflect an iterative, integrated testing approach enabled by digital twins and digital threads to support delivery of minimum viable products.

DoD Response: PARTIALLY CONCUR – Existing Department of the Navy (DON) Test and Evaluation policy, included in the DON's acquisition instruction, Secretary of the Navy Instruction (SECNAVINST) 5000.2G, currently directs programs to conduct Capabilities Based Test and Evaluation (CBTE), which is based on the warfighter tasks and conditions identified in the Mission Based Test Design, and maps all testing and identified deficiencies to mission capabilities and mission impact. CBTE institutionalizes integrated testing. SECNAVINST 5000.2G also provides a requirement for building digital models and sharing data but does not specify digital threads or digital twins. Once the Department of Defense issues specific language to describe these activities in its respective acquisition policy documents, DON will incorporate those changes in the next update to SECNAVINST 5000.2G.

Recommendation 12: The Secretary of the Navy should revise the military department's weapon system test and evaluation policy to fully reflect leading practices for product development by requiring developmental and operational testers to develop and implement test strategies and test plans that incorporate end user agreements that detail a process for obtaining ongoing user input and feedback for the system under test.

DoD Response: DO NOT CONCUR – Existing Department of the Navy acquisition policy states, "the program collaborates with Resource Sponsor/Requirements Office/Fleet/Force/OTA to identify applicable mission tasks, measures and conditions, as derived from applicable documents (including Required Operational Capabilities (ROC) and Projected Operational Environment (POE), Naval Tactical Tasks, Marine Corps Tasks, Coast Guard Tasks, etc.) for the System Under Test." The report is unclear in establishing what additional end user coordination would accelerate delivery of warfighting capability.

Recommendation 13: The Secretary of War should ensure that the Under Secretary of War for Acquisition and Sustainment revise its weapon system acquisition policies to require that such acquisition programs obtain and incorporate input from developmental and operational testers in the development of acquisition strategies, including on testing-related topics such as the use of and access to digital twins and digital threads.

DoD Response: The Office of the Under Secretary of War for Acquisition and Sustainment concurs with this recommendation.

Appendix III: GAO Contact and Staff Acknowledgments

GAO Contact

Shelby S. Oakley, oakleys@gao.gov

Staff Acknowledgments

In addition to the contact named above, Christopher R. Durbin (Assistant Director), Ashley Rawson (Analyst-in-Charge), Pete Anderson, John Armstrong, Matthew T. Crosby, Lori Fields, Scott Hepler, Kathryn Long, and Benjamin Nilson were key contributors to this report.

GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through our website. Each weekday afternoon, GAO posts on its website newly released reports, testimony, and correspondence. You can also subscribe to GAO's email updates to receive notification of newly posted products.

Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, https://www.gao.gov/ordering.htm.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or
TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

Connect with GAO

Connect with GAO on X, LinkedIn, Instagram, and YouTube.
Subscribe to our Email Updates. Listen to our Podcasts.
Visit GAO on the web at https://www.gao.gov.

To Report Fraud, Waste, and Abuse in Federal Programs

Contact FraudNet:

Website: https://www.gao.gov/about/what-gao-does/fraudnet

Automated answering system: (800) 424-5454

Media Relations

Sarah Kaczmarek, Managing Director, Media@gao.gov

Congressional Relations

A. Nicole Clowers, Managing Director, CongRel@gao.gov

General Inquiries

https://www.gao.gov/about/contact-us