



June 2015

DOD OPERATIONAL TESTING

Oversight Has
Resulted in Few
Significant Disputes
and Limited Program
Cost and Schedule
Increases

Accessible Version

GAO Highlights

Highlights of [GAO-15-503](#), a report to congressional committees

Why GAO Did This Study

DOD conducts extensive operational testing and evaluation of its military systems prior to full-rate production and fielding. DOT&E plays an integral role in operational test and evaluation by issuing policy and procedures, overseeing operational test planning, and independently evaluating and reporting test results. At times, DOT&E and acquisition programs may disagree about what is needed to adequately demonstrate operational capability, which sometimes may affect programs' cost or schedule.

The Joint Explanatory Statement to Accompany the National Defense Authorization Act for Fiscal Year 2015 directed GAO to review DOT&E's oversight activities. This report examines (1) the extent to which DOD acquisition programs have had significant disputes, if any, with DOT&E over operational testing, and (2) the circumstances and impact of identified disputes. GAO evaluated documentation and interviewed officials from DOT&E, other DOD test organizations, and the acquisition community. GAO also conducted 10 case studies from among 42 programs identified by military service officials as having had significant disputes with DOT&E. GAO analyzed information received from acquisition and testing officials to verify the merits and degree of those disputes. Based on this assessment, GAO selected case studies that were representative of the most significant disputes identified across the military services.

What GAO Recommends

GAO is not making any recommendations in this report.

View [GAO-15-503](#). For more information, contact Michael Sullivan at (202) 512-4841 or sullivanm@gao.gov.

June 2015

DOD OPERATIONAL TESTING

Oversight Has Resulted in Few Significant Disputes and Limited Program Cost and Schedule Increases

What GAO Found

The Director, Operational Test and Evaluation (DOT&E) provided oversight for 454 Department of Defense (DOD) acquisition programs from fiscal years 2010-2014. Military service officials identified 42 programs from that period that they believed had significant disputes with DOT&E over operational testing—that is, disputes that may have led to cost and schedule impacts for programs. Operational testing is intended to evaluate a system's capability in realistic combat conditions before full-rate production or full deployment. Acquisition programs and DOT&E have different objectives and incentives, which can potentially fuel tension between the two over what is needed to accomplish operational testing for programs. According to military service officials, the tension is generally manageable and differences usually are resolved in a reasonable and timely manner, with modest adjustments often required in the course of developing and executing a test approach. However, sometimes differences about operational testing requirements, methods, costs, or results develop into significant disputes and are more difficult to resolve. Acquisition and test officials from the military services identified only a limited number of cases—less than 10 percent of programs receiving DOT&E operational test oversight since fiscal year 2010—that they believed had experienced significant operational testing disputes with DOT&E. Officials noted that although these disputes can require additional time and effort to work through, they generally get resolved.

In an in-depth review of 10 case studies selected from among the 42 programs with significant disputes, GAO identified a variety of factors that contributed to disputes between the acquisition programs and DOT&E, but only a few cases that involved considerable cost or schedule impacts. Key factors involved the adequacy of proposed testing and differences over test requirements, assets, and the reporting of test results. In general, GAO found that DOT&E had valid and substantive concerns about operational test and evaluation for each of the 10 cases reviewed. However, military service officials indicated to GAO that testing advocated by DOT&E was, in some instances, beyond what they believed was necessary and lacked consideration for programs' test resource limitations. Many of the disputes GAO reviewed were, or are expected to be, resolved in DOT&E's favor with limited cost and schedule impacts to the programs. In a few cases, military service officials acknowledged that benefits were achieved from resolving the disputes, such as a reduction in the scope of operational testing and better understanding of system requirements. Resolution of disputes for three programs—DDG-51 Flight III Destroyer, F-35 Joint Strike Fighter, and CVN 78 aircraft carrier—had considerable potential or realized cost or schedule effects and required formal involvement from senior DOD leadership. For the first two programs, hundreds of millions of dollars in additional costs were associated with resolving their disputes. For CVN 78, the dispute—which remains unresolved—involves the Navy's carrier deployment schedule and whether survivability testing will be deferred by several years. For the other seven case study programs that GAO reviewed, the cost and schedule effects tied to dispute resolution were more limited, and in some instances, not related to operational testing requirements.

Contents

Letter	1
Background	2
Inherent Tension Exists in Operational Test Oversight, but Few Programs Experienced Significant Disputes with DOT&E	7
A Variety of Factors Contributed to Disputes We Reviewed, but Only a Few Cases Had Significant Cost or Schedule Impacts	11
Agency Comments	27
Appendix I: Objectives, Scope, and Methodology	29
Appendix II: GAO Contact and Staff Acknowledgments	33
GAO Contact	33
Staff Acknowledgments	33
Appendix III: Accessible Data	34
Table	
Table 1: Operational Test Dispute Factors Identified for 10 GAO Case Studies	14
Data Table for Figure 1: Programs on the Director, Operational Test and Evaluation (DOT&E) Oversight List, Fiscal Years 2005-2014	34
Figures	
Figure 1: Programs on the Director, Operational Test and Evaluation (DOT&E) Oversight List, Fiscal Years 2005-2014	5
Figure 2: Director, Operational Test and Evaluation (DOT&E) Typical Involvement in the DOD Acquisition Cycle	6

Abbreviations

DOD	Department of Defense
DOT&E	Director, Operational Test and Evaluation
FOT&E	Follow-on operational test and evaluation
FSST	Full ship shock trial
IOT&E	Initial operational test and evaluation
LFT&E	Live fire test and evaluation
SDTS	Self-defense test ship
TEMP	Test and evaluation master plan
USD (AT&L)	Under Secretary of Defense for Acquisition, Technology, and Logistics

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.



June 2, 2015

Congressional Committees

The Department of Defense (DOD) develops and acquires some of the most advanced military systems in the world. The development of these systems often involves new technologies, complex designs, and the integration of multiple subsystems and components. DOD conducts extensive operational testing and evaluation of the systems prior to full-rate production and fielding to ensure that warfighters have an understanding of the capabilities and limitations of each system. As authorized under certain sections of Title 10 of the United States Code, the Director, Operational Test and Evaluation (DOT&E) plays an integral role in operational test and evaluation by issuing policy and procedures, overseeing operational test planning, independently evaluating and reporting on test results, and advising senior DOD decision-makers and Congress on the operational capabilities of systems.¹ At times, DOT&E and acquisition programs or military service test and oversight organizations may disagree about what is required to adequately demonstrate operational capability. These disagreements may have cost, schedule, or performance implications for acquisition programs. The Joint Explanatory Statement to Accompany the National Defense Authorization Act for Fiscal Year 2015 directed GAO to review DOT&E's oversight activities and any potential impact they may have on acquisition programs.² This report examines (1) the extent to which there have been any significant disputes between DOT&E and DOD acquisition programs over operational testing, and (2) the circumstances and impact of identified operational test-related disputes.

To conduct this work, we reviewed documentation from and interviewed relevant DOD acquisition and test officials. To determine the extent to which significant disputes between DOT&E and acquisition programs may have occurred, we reviewed DOD documentation and obtained formal input from senior military service acquisition and test officials and operational test agencies within the military services. Based on their

¹ 10 U.S.C. §§ 139, 2399.

² 160 Cong. Rec. H8671, H8703 (Dec. 4, 2014).

familiarity with the operational test process with DOT&E, military service officials identified 42 programs that had significant disputes with DOT&E between fiscal years 2010 and 2014, involving one or more of the following: (1) delay in DOT&E approval or a disapproval of test planning documentation, (2) disagreement over the scope, design, or assets needed for testing, (3) increases in operational test costs that were believed to be unwarranted, and (4) disagreement about DOT&E's characterization of test outcomes in its formal reporting to DOD and congressional defense committees. To assess the circumstances and impact of identified disputes, we completed 10 in-depth reviews (case studies) from the 42 programs that military service officials identified as having been associated with some of the most significant disputes within each military service in recent years. To complete these case studies, we reviewed program and testing information, and interviewed appropriate officials from the acquisition and testing communities, as well as DOT&E officials.

We conducted this performance audit from July 2014 to June 2015 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

Test and evaluation activities are an integral part of developing and producing weapon systems, as they provide knowledge of a system's capabilities and limitations as it matures and is eventually delivered for use by the warfighter. DOD divides its testing activities into three categories: developmental, operational, and live fire. Developmental testing, which is conducted by contractors, university and government labs, and various DOD organizations, is intended to provide feedback on the progress of a system's design process and its combat capability as it advances toward initial production or deployment. Operational test and evaluation is intended to evaluate a system's effectiveness and suitability under realistic combat conditions before full-rate production or deployment occurs. DOD defines operational effectiveness as the overall degree of mission accomplishment of a system when used by representative personnel (e.g., warfighters) in the environment planned or expected for operational employment of the system considering organization, training, doctrine, tactics, survivability or operational security, vulnerability, and threat. Operational suitability defines the

degree in which a system can be satisfactorily placed in field use, with consideration given to its reliability, transportability, interoperability, and safety, among other attributes. Under live fire test and evaluation, when applicable, survivability is a measure of a system's vulnerability to munitions likely to be encountered in combat and lethality measures a system's ability to combat intended targets.³ Operational testing is managed by the various military test organizations representing the customers, such as combat units that will use the weapons. Each of the four military services—Air Force, Army, Marine Corps, and Navy—has its own operational test agency to plan and execute testing.

In 1983, Congress established DOT&E to coordinate, monitor, and evaluate operational testing of major weapon systems. As part of the Office of the Secretary of Defense, DOT&E is separate from the acquisition and test communities within the military services and other defense agencies. This enables DOT&E to provide the Secretary of Defense and Congress with an independent perspective on operational or live fire testing activities and results for DOD acquisition programs. DOT&E serves as the principal adviser on operational test and evaluation in DOD and bears several key responsibilities, which include:

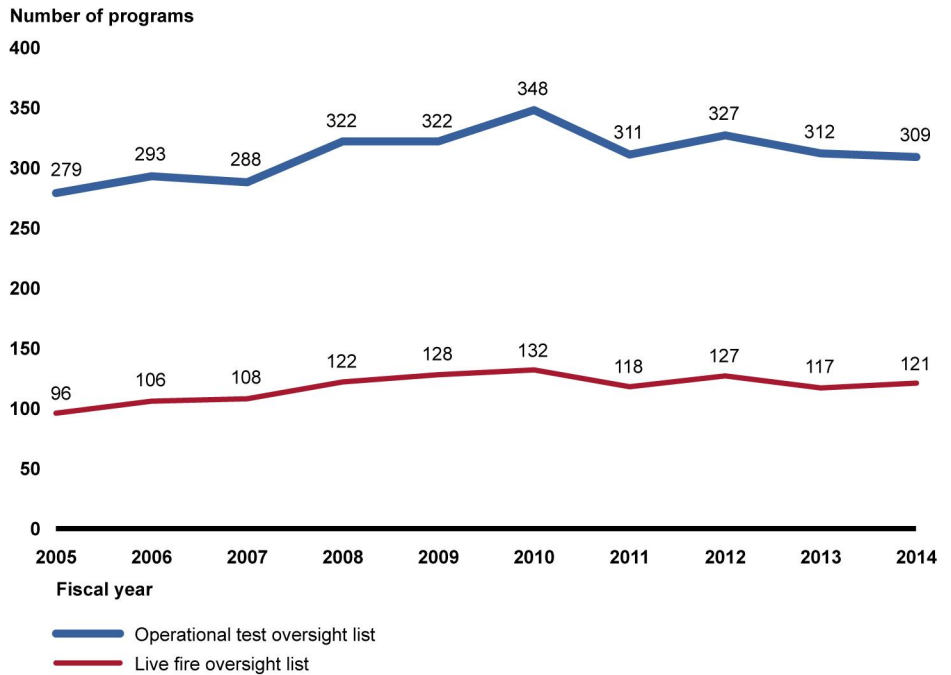
- providing policy and guidance to DOD leadership on operational and live fire test and evaluation;
- monitoring and reviewing all operational and live fire test and evaluation events;
- approving test and evaluation master plans (TEMPs), operational test plans, and live fire test plans for all programs receiving oversight;
- reporting to the Secretary of Defense and congressional defense committees on programs generally before a full-rate production decision regarding (1) the adequacy of operational and live fire test and evaluation, and (2) operational effectiveness and suitability for combat; and
- reporting annually to the Secretary of Defense and Congress on all operational and live fire test and evaluation results from the preceding fiscal year.

³ DOD, Defense Acquisition Guidebook (Sept. 2013).

By law, DOT&E is responsible for overseeing all major defense acquisition programs, as well as any other acquisition programs it determines should be designated for oversight.⁴ The number of programs receiving operational test and evaluation oversight from DOT&E increased considerably between 2005 and 2010, from 279 to 348 programs, but since that time has declined, averaging 315 programs per year. Programs under DOT&E live fire test and evaluation oversight increased in a similar way over the past decade, averaging about 121 programs per year since 2011. DOT&E may oversee operational testing or live fire testing or both, depending on the circumstances of each program. Figure 1 shows the number of programs receiving DOT&E oversight annually over the last decade.

⁴ Non-major programs typically receive DOT&E oversight if they require joint or multi-service testing, have a close relationship to or are a key component of a major program, are an existing system undergoing major modification, or are of special interest—often based on input or action from Congress.

Figure 1: Programs on the Director, Operational Test and Evaluation (DOT&E) Oversight List, Fiscal Years 2005-2014

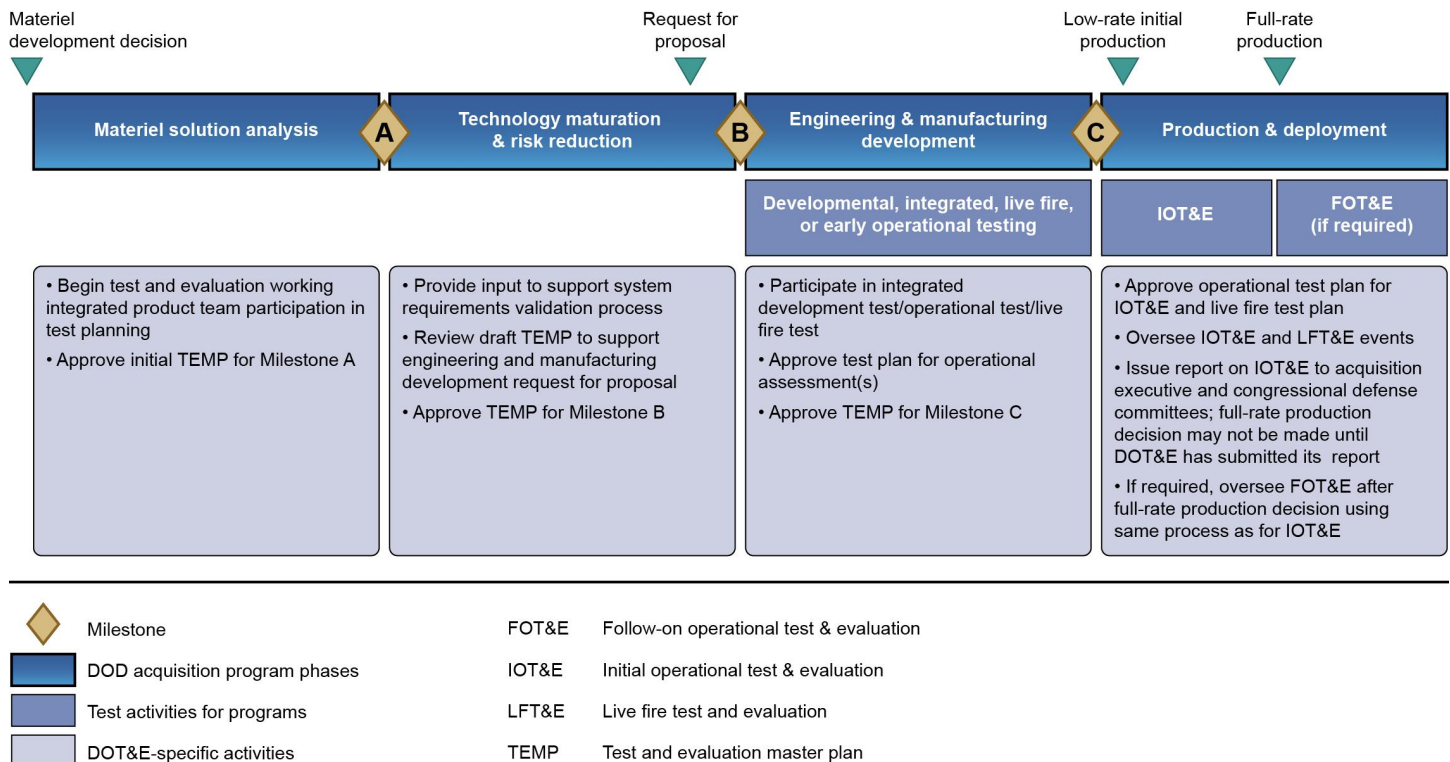


Source: GAO analysis of DOT&E annual reports for fiscal years 2005-2014. | GAO-15-503

Note: Programs can be on DOT&E oversight for operational testing, or live fire testing, or both.

Generally, programs are added to the oversight list when they formally enter the acquisition process, and DOT&E oversight continues through key acquisition milestones to full-rate production approval. Figure 2 illustrates the acquisition process, test phases, and DOT&E's involvement in oversight.

Figure 2: Director, Operational Test and Evaluation (DOT&E) Typical Involvement in the DOD Acquisition Cycle



Source: GAO analysis of DOD Instruction 5000.02 acquisition guidance. | GAO-15-503

During their planning phase, DOD acquisition programs establish a test and evaluation working integrated product team that is made up of acquisition and test stakeholders, including DOT&E representatives. The main focus of this team is developing the TEMP, which provides a framework and plan for what developmental and operational testing will be conducted, as well as test resources needed, and how the major test events and test phases for a program link together. The TEMP also identifies criteria to be used to test and evaluate the system. The TEMP is required for key program milestone reviews, such as Milestone B and Milestone C, and as required by DOD policy, must be approved by DOT&E and several other DOD and military service organizations.⁵

⁵ DOD Instruction 5000.02, *Operation of the Defense Acquisition System*, Enclosure 5 (Jan. 2015).

During the later stages of system development (engineering and manufacturing development phase) and before a program has production units of its system available for testing, one or more operational assessments may be conducted. These assessments usually are completed by the designated operational test agency in accordance with a test plan approved by DOT&E, and the results can be used to inform the Milestone C initial production decision for a program. Following Milestone C, programs develop an operational test plan to support the initial operational test and evaluation (IOT&E) of the system. This test plan is expected to provide a detailed scope and methodology for conducting operational and live fire test events. By law, DOT&E approval of the operational test plan is required for programs on the oversight list, and the operational test agency conducts testing in accordance with the DOT&E-approved plan.⁶ Representative users (e.g., the warfighters) and production-representative units (e.g., systems from low-rate initial production) are used for IOT&E to determine if a system is operationally effective and operationally suitable for its mission. DOT&E formally monitors the IOT&E event, and reports the results of its evaluation to the Secretary of Defense and Congress, as required, in support of a full-rate production decision.

Inherent Tension Exists in Operational Test Oversight, but Few Programs Experienced Significant Disputes with DOT&E

Different objectives and incentives exist between the acquisition and testing communities, which can potentially fuel tension over what is needed to accomplish operational testing for programs. According to DOD officials, differences usually are resolved in a reasonable and timely manner, with modest adjustments often required in the course of developing and executing a test approach. However, sometimes acquisition and test officials have differences about operational testing requirements, methods, costs, or results that develop into significant disputes that are more difficult to resolve. Acquisition and test officials in the military services identified a small number of cases where significant disputes occurred among programs receiving DOT&E oversight during fiscal years 2010 through 2014. Of the 454 programs on DOT&E's oversight list during that period, officials identified 42 programs—less than

⁶10 U.S.C. 2399 (b) requires that major defense acquisition programs, designated as such by DOT&E, receive approval for operational test plans before conducting operational testing. DOD Instruction 5000.02 requires DOT&E approval for test and evaluation master plans at key acquisition milestone decisions.

10 percent—that had significant operational test disputes with DOT&E.⁷ According to military service and DOT&E officials, while these disputes can require additional time and effort to work through, they generally do get resolved.

Inherent Tension in Operational Test Oversight

Acquisition and test officials from the military services stated that, in general, DOT&E's execution of its oversight authorities provided valuable input and support to acquisition programs and to the entire DOD operational test enterprise. However, at times in the past, some elements of the DOD acquisition community, such as program offices, program executive offices, or senior acquisition executive offices, have expressed concerns that the test community's approach to testing imposes undue requirements on programs. In response to concerns like these, in 2011 the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD (AT&L)) chartered an independent assessment of the developmental and operational test communities' approach to testing. The assessment found no significant evidence that the testing community typically drives unplanned requirements, cost, or schedule into programs.⁸ The assessment, however, acknowledged that tension exists between programs and the test community, but noted it can be mitigated through early and objective communication of issues, and when necessary, through involving senior DOD leadership.

Tension exists between the acquisition and testing communities, in part, because they have somewhat different objectives and perspectives regarding the role of operational testing. Acquisition managers are motivated largely by their programs' cost, schedule, and performance objectives, particularly once they have an approved program baseline that formalizes those objectives for system development. DOT&E is focused on ensuring the operational effectiveness and suitability of systems is adequately evaluated. As program managers work to preserve their program goals, they may see test reductions or delays as a reasonable option for offsetting cost and schedule growth encountered during system

⁷ The total number of acquisition programs was determined through our analysis of oversight lists from DOT&E's annual reports for fiscal years 2010 through 2014. Many programs were on the oversight list for multiple years during this period.

⁸ Office of the Secretary of Defense, *Memorandum: Test and Evaluation of Defense Programs* (Washington, D.C.: Jun. 3, 2011).

development. We have previously found that compromises in test approaches and resources are more readily made in deference to other priorities, such as preserving program cost goals. Postponing difficult tests or limiting open communication about test results can help a program avoid unwanted scrutiny because tests against criteria can reveal shortfalls, which may call into question whether a program should proceed as planned.⁹

DOT&E's approval authority for core test documents—particularly TEMP's and operational test plans—also contributes to the tension. Approval can be withheld until these documents demonstrate adequate means to evaluate operational effectiveness and suitability. TEMP's must also show that sufficient resources have been dedicated to the operational test program. Costs for operational tests are predominantly borne by the programs, which creates another source of tension. Although operational testing typically represents a relatively small amount of the total program cost to develop and produce a system, this cost can be significant in the years during which operational testing events occur. In 2011, DOT&E assessed 78 recent acquisition programs and found the average marginal cost of operational test and evaluation to be about 1 percent of total program acquisition costs.¹⁰ Additionally, programs may lack the funding and contract flexibility to accommodate discovery and respond to changes in testing needs during program execution. Aside from the cost, the fact that operational testing occurs largely in the later stages of a program—when overall research, development, test, and evaluation funds generally are more limited—may also create significant challenges if test changes are needed. This may be particularly true when production has already begun. In those cases, additional costs can stem both from increased testing and from production delays, as well as any potential retrofitting required for the systems already produced.

DOT&E officials stated that test plan approval is an iterative process, and considerable time and resources are spent by DOT&E, programs, and operational test agencies to finalize these documents. Overall, military service officials we interviewed indicated that the give-and-take between programs and DOT&E in developing test plans is generally manageable,

⁹ GAO, *Best Practices: A More Constructive Test Approach Is Key to Better Weapon System Outcomes*, [GAO/NSIAD-00-199](#) (Washington, D.C.: Jul. 2000).

¹⁰ Director, Operational Test and Evaluation, *FY 2011 Annual Report* (Dec. 2011).

but voiced concern over how long it sometimes takes to receive final DOT&E approval. Officials indicated that approval can be challenging, in part, because good relationships and communications between DOT&E and program and test officials from the military services sometimes are lacking during the test plan development process. For example, some officials voiced frustration about mixed messages sometimes received from DOT&E's working-level officials—action officers—and senior officials, where action officers may have agreed to test plans at a working level, but senior DOT&E officials rejected the plans.

Program and test officials noted that tension can be amplified when a program has spent substantial time working on test plans, and is approaching a major acquisition milestone or scheduled operational test event, but has not yet received DOT&E approval. In a limited number of instances, DOT&E has formally disapproved program TEMP's and operational test plans. Specifically, DOT&E annual reports for fiscal years 2010 through 2014 show that 245 TEMP's and 375 operational test plans were approved, and only 14 test documents were disapproved.¹¹ Several TEMP's or operational test plans that were disapproved were later resubmitted with changes and approved. Some cases of test plan disapproval were the result of systems not being ready for operational testing. DOT&E officials noted, however, that disapproval of test plans does not directly indicate a program had a significant dispute with DOT&E. For example, one TEMP was disapproved because the program mistakenly submitted a prior version of the plan for approval that did not have updates that had been agreed to by the program and DOT&E.

Few Significant Operational Test Disputes Were Identified

In the absence of any definitive indicators being found during our review that could be used to identify cases of significant operational testing disputes, we asked acquisition and test officials from the military services to identify programs that had significant disputes with DOT&E from among the 454 programs on DOT&E's oversight list in fiscal years 2010 through 2014. In response, officials identified 42 programs—less than 10 percent of the total—that they believed had significant disputes with DOT&E. Significance, while subjective, was judged by the officials based

¹¹ The number of approved TEMP's includes a limited number of test and evaluation strategies—precursors to TEMP's—from fiscal year 2010 because DOT&E did not distinguish between the two in that year's report. Test and evaluation strategy approvals ranged from 1 to 6 per year based on fiscal year 2011-2014 reports.

on their familiarity with what is typical for the operational test process with DOT&E and what programs encountered problems that they believed went beyond that norm. Officials identified programs that they believed had significant disputes with DOT&E related to operational or live fire testing oversight involving one or more of the following: (1) substantial delay in DOT&E approval, or a disapproval, of test planning documentation; (2) significant disagreement over test asset needs or test scope; (3) considerable increases in operational test costs that were believed to be unwarranted; and (4) disagreements about DOT&E's characterization of test outcomes in its formal reporting to the Secretary of Defense and congressional defense committees. For many of the disputes identified by military service acquisition and test officials, opinions varied as to whether they were significant or simply indicative of the usual back-and-forth that occurs when planning for and executing operational testing. Military service and DOT&E officials, however, noted that although disputes can require additional time and effort to work through, they generally do get resolved.

A Variety of Factors Contributed to Disputes We Reviewed, but Only a Few Cases Had Significant Cost or Schedule Impacts

To gain a better understanding of the circumstances of disputes and assess the merits of disputes for cases where opinions varied, we selected and performed an in-depth review of 10 cases from among the 42 programs that had disputes. We believe these cases were among the most significant within each of the military services. The 10 cases had a variety of factors that contributed to disputes between DOT&E and military service officials over operational testing, but only a few had considerable cost or schedule impacts. The factors typically involved the adequacy of proposed testing and differences over test requirements, assets, and the reporting of test results. In general, we found that DOT&E had valid and substantive operational test-related concerns for each program reviewed. On the other hand, military service officials we interviewed contended that in some instances testing advocated by DOT&E was in excess of what was needed to determine the operational effectiveness or suitability of systems or unrealistic given the test resource limitations for programs.

Most of the disputes we reviewed had been, or are expected to be, resolved in favor of DOT&E's concerns, and with limited cost and schedule impacts to the programs. In a few of these cases, military service officials acknowledged that benefits were achieved from resolving the disputes, such as a reduction in the scope of operational testing and better understanding of system requirements. However, resolution of disputes for three programs—DDG-51 Flight III Destroyer, F-35 Joint

Strike Fighter, and CVN 78 aircraft carrier—had considerable potential cost or schedule effects that required formal involvement from senior DOD leadership. For the first two programs, hundreds of millions of dollars in additional costs are associated with resolving their disputes. For CVN 78, the dispute, which remains unresolved, has ramifications for the Navy’s carrier deployment schedule and whether a key test to assess the survivability of the carrier will be deferred by several years.

Several Factors Contributed to Operational Testing Disputes for Case Study Programs

We identified five primary factors that contributed to operational test disputes with DOT&E among our 10 cases. These factors collectively revolved around the adequacy of the scope, design, and execution of operational testing for programs and what is needed to ensure a system is tested in a manner that represents its intended operational environment. The five factors include (1) poorly-defined system requirements, (2) the need to address operational threats and environments, (3) insufficient test assets, (4) live fire test issues, and (5) disagreements over the reporting of test results.

- System performance requirements—approved by the military services and DOD’s Joint Requirements Oversight Council—played a role in the disputes for five programs we reviewed. The associated issues we found involved the ability to adequately test and evaluate requirements, requirements that were not believed to reflect how a system would be used in combat, and requirements not being operationally tested as planned. For example, DOT&E determined for several of the programs we reviewed that program performance requirements were not directly linked to measures of mission success, and therefore not adequate as a basis for testing the system’s effectiveness and suitability.
- Testing against relevant operational threats or in relevant operational environments was a basis of disputes for three of our case studies. They included issues with ensuring current operational threats were accounted for in testing and with making certain that specified test locations reflect the intended operational environment for a system. As an example, in the case of an Army self-propelled howitzer program, DOT&E identified that the operational environment proposed by the Army was inconsistent with the way the system would likely be employed in the field, which led to a disagreement related to the survivability of the system.
- Test asset differences were tied to disputes for three of our case studies. Issues related to this factor stemmed from programs and

DOT&E having different perspectives on the type of test assets needed to demonstrate operational effectiveness and suitability, or the need for test assets not planned for by programs. The most substantial example of this in our cases involved a disagreement between DOT&E and the Navy over whether an unmanned self-defense test ship was needed to demonstrate operational effectiveness and suitability for a Navy destroyer and its major subsystems.

- Live fire testing issues, which factored into disputes for three cases, dealt with disagreements over the timing or extensiveness of testing, including the use of existing data to support evaluations of operational survivability. The most significant case tied to this factor involves a disagreement over when and on what carrier a full ship shock trial—a live fire test of the survivability of the new aircraft carrier and its subsystems—will be completed.
- Disagreement with DOT&E’s characterization of test results in its operational test reports contributed to the disputes with two programs. In particular, both programs’ officials took exception to the manner in which DOT&E discussed test results, noting that they believed DOT&E obscured the fact that the systems were demonstrated through testing to meet their requirements. DOT&E officials emphasized that the system performance against requirements was clearly stated in reports, but that they also are responsible for characterizing any limitations to testing performed or limitations to system capability identified.

Table 1 provides our assessment of which factors contributed to the disputes for each of the 10 case studies we completed.

Table 1: Operational Test Dispute Factors Identified for 10 GAO Case Studies

Program	System requirements	Operational threats and environments	Test assets	Live fire test issues	Reporting of test results
CVN 78 <i>Gerald R. Ford</i> Class Aircraft Carrier				X	
DDG-51 Flight III Destroyer / AN/SPY-6 Radar / Aegis Modernization			X		
DOD Automated Biometrics Identification System	X				
Enhanced Combat Helmet				X	X
F-35 Joint Strike Fighter / Electronic Warfare Infrastructure Improvement Program		X	X		
Ground / Air Task Oriented Radar	X		X		
Joint Assault Bridge		X		X	
P-8A Poseidon Multi-Mission Maritime Aircraft	X				X
Paladin Integrated Management	X	X			
Three-Dimensional Expeditionary Long-Range Radar	X				

Source: GAO analysis; DOD interviews and documentation. | GAO-15-503

Most Disputes Were Resolved without Significant Cost or Schedule Impacts

In reviewing 10 case studies, we found that DOT&E raised legitimate concerns about the ability to adequately operationally test the systems and evaluate their effectiveness and suitability. However, we also recognize the real concerns voiced by military officials about the difficulties in reaching agreement with DOT&E for these cases and the potential or realized cost and schedule consequences for programs. Although the true cost of overcoming a significant disagreement is not easily measured, in general, we found that many of the disputes between programs and DOT&E outlined in our case studies appear to have been, or are expected to be, resolved with relatively limited effects on program cost or schedule. There were, however, three case studies that had disputes with substantial cost or schedule implications that did, or may, require decisions from top DOD leadership—the Deputy Secretary of Defense or the Secretary of Defense—to resolve. DOT&E officials stated they believe these are the only cases among the 454 programs on DOT&E oversight since fiscal year 2010 that required this type of senior-level involvement in order to resolve a dispute. Resolving these disputes carried substantial impacts, although not always borne by the program

offices themselves. For two of the programs—the DDG-51 Flight III Destroyer and Joint Strike Fighter—hundreds of millions of dollars in additional costs were associated with their dispute resolutions.¹² For the third program—CVN 78—the dispute, which has yet to be resolved, has ramifications for the Navy’s carrier deployment schedule and whether survivability testing, which is intended to identify potential vulnerabilities to the carrier and reduce risk to sailors, will be deferred by several years. The following profiles provide details on the disputes, resolution, and associated impacts for these three case studies.

¹² The DDG-51 Flight III Destroyer, AN/SPY-6 Radar, and Aegis Modernization programs, which will be integrated into a unified weapon system, are part of the same dispute, so they were treated as a single case for our analysis.

CVN 78 Gerald R. Ford Class Aircraft Carrier

The Director, Operational Test and Evaluation (DOT&E) has been engaged with the Navy in a dispute over whether to conduct the full ship shock trial (FSST) on CVN 78—the first of the new class of nuclear-powered aircraft carriers—as previously agreed to in the program’s alternative Live Fire Test and Evaluation Management Plan signed by the Navy and DOT&E in 2007, or to defer it to the follow-on ship (CVN 79) as the Navy decided in 2011 due to technical, schedule, and budgetary concerns. FSST is a test that employs an underwater charge at a certain distance from the carrier to identify survivability issues for the ship and its key systems. Early discovery of issues may then be used to implement fixes while follow-on carriers are still being built to assure their survivability and reduce risk to sailors. The Navy believes lessons learned from FSSTs on other ships, when combined with shock testing being performed on individual ship components and equipment, reduce the need to complete FSST on CVN 78. DOT&E provided memoranda to the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD (AT&L)) and the Navy that documented the findings from previous FSST events for other ships and concluded that those results made component-level testing and past FSST results insufficient to assess survivability of the new carrier class.

Impact: Completing FSST on CVN 78 could delay deployment of the carrier 1-6 months based on current estimates. The Navy has stated that any deployment delay would further delay returning its fleet size to the congressionally-mandated 11 carriers. DOT&E has emphasized that, regardless of any change to FSST, a carrier fleet size shortfall will exist for at least 5 years—the shortfall has existed since the CVN 65 carrier was decommissioned in 2012—and the 5- to 7-year delay associated with deferring the test to CVN 79 would reduce the potential to discover survivability problems early and fix them. In addition, as we recently found in a review of the carrier program, CVN 78 has faced construction challenges and issues with key technologies that increase the likelihood the carrier will not deploy as scheduled or will deploy without fully tested systems.¹

Resolution status: DOT&E and the Navy have been unable to resolve this dispute. In May 2015, the Navy revised its position on the FSST, presenting a plan to USD (AT&L) to conduct the test on CVN 78, but not until sometime after the ship’s first deployment. The Navy stated this would preserve the ability to deploy CVN 78 and meet the 11-carrier fleet requirement at the earliest opportunity. DOT&E disagreed with the Navy’s new plan to complete FSST after deployment and reiterated that completing testing before deployment is the only way many shock-related survivability issues can be found and addressed before the ship and crew deploy into an active theater of operations. DOD leadership is expected to resolve this dispute later in 2015.

Source: GAO analysis; DOD interviews and documentation. | GAO-15-503

¹ GAO, *Ford-Class Aircraft Carrier: Congress Should Consider Revising Cost Cap Legislation to Include All Construction Costs*, GAO-15-22 (Washington, D.C.: Nov. 2014).

DDG-51 Flight III Destroyer / AN/SPY-6 Radar / Aegis Modernization

The Navy and the Director, Operational Test and Evaluation (DOT&E) have an ongoing dispute over the need to use an unmanned self-defense test ship (SDTS) to accomplish operational testing of the next Aegis combat system and AN/SPY-6 radar on the DDG 51 Flight III Destroyer—a multi-mission ship designed to defend against air, surface, and subsurface threats. DOT&E expects these systems to be tested together to ensure operationally realistic testing and an end-to-end assessment of the ship's capability; an approach which has been used for other Navy surface ship programs. DOT&E disapproved test and evaluation master plans for the Aegis and AN/SPY-6 programs because the Navy did not include the use of the SDTS. DOT&E's analysis concluded that a SDTS, equipped with the Aegis and AN/SPY-6 systems, is needed for close-in live fire testing against most classes of anti-ship cruise missile threats, including supersonic, maneuvering threats—a manned ship cannot be used because of safety concerns. DOT&E also emphasized that past testing using an unmanned SDTS led to the discovery of combat system deficiencies that could not have been found by using constrained testing approaches against manned ships. Navy officials believe their test approach, which relies on collecting data from multiple sources—live fire end-to-end testing of selected targets on a tactical manned ship, limited missile intercept testing using the existing SDTS, and land-based test sites—achieves a better balance between cost and risk. DOT&E officials emphasized that the Navy's test approach will not provide the data needed to validate modeling and simulation and is insufficient to demonstrate ship self-defense capabilities and survivability against operationally realistic threats. In particular, DOT&E stated the proposed live fire testing on the tactical manned ship and land-based testing are constrained considerably because of safety restrictions, and the Navy's proposed missile intercept testing using the existing SDTS does not provide the needed data because it uses different combat and launching systems than those intended for the DDG-51 Flight III Destroyer.

Impact: Preliminary estimates suggest the additional cost of using SDTS for operational testing would be \$320-\$470 million, with DOT&E officials noting the actual cost is likely to be somewhere in the middle of that range. The Navy has not determined the difference in total test cost if SDTS is used versus some alternative approach, but has estimated the cost of the modeling and simulation suite to support testing at \$86.7 million over the next 5 years. DOT&E estimates that about \$230 million of the test cost with SDTS could potentially be recovered by the Navy if the systems installed on the SDTS are removed after testing and integrated on a future DDG-51 Flight III ship.

Resolution status: DOT&E and the Navy have not resolved this dispute. The Office of Cost Analysis and Performance Evaluation within the Office of the Secretary of Defense is expected to complete an analysis in June 2015 on the cost to upgrade an existing SDTS, which is intended to inform a decision by the Deputy Secretary of Defense on whether a SDTS will be used for initial operational test and evaluation.

Source: GAO analysis; DOD interviews and documentation. | GAO-15-503

F-35 Joint Strike Fighter and the Electronic Warfare Infrastructure Improvement Program

In early 2012, the Director, Operational Test and Evaluation (DOT&E) identified shortfalls in DOD's electronic warfare test capabilities that posed problems for operationally testing the Joint Strike Fighter, the next generation fighter aircraft. Specifically, a threat assessment report outlined current threats that raised questions regarding the performance of the Joint Strike Fighter aircraft and other systems when employed against those threats. DOT&E indicated that additional investment was needed to upgrade outdoor test range assets, anechoic chambers (a room designed to completely absorb reflections of electromagnetic waves), and electronic warfare programming labs in order to test against updated threats as required. Joint Strike Fighter officials agreed that the aircraft should be tested against current threats, but emphasized that the program should not have to fund these test infrastructure improvements. To assess the issue further, the Office of the Secretary of Defense commissioned a study of electronic warfare test infrastructure needs.

Impact: The Office of the Secretary of Defense study validated DOT&E's concerns, concluding that test infrastructure improvements were needed to support testing of the Joint Strike Fighter and a number of other systems being developed.

Resolution status: In response to the study, the Secretary of Defense signed a Resource Management Decision in September 2012 that established the Electronic Warfare Infrastructure Improvement Program to acquire and upgrade electronic warfare test capabilities that are intended to support operational testing for the Joint Strike Fighter and other systems. The decision provided about \$491 million outside of the Joint Strike Fighter program funding for the Electronic Warfare Infrastructure Improvement Program. Plans for the program include procuring 22 emitters to support the full range of testing needs. Joint Strike Fighter program officials said they expect to begin testing with whatever assets are available to meet the test schedule.

Source: GAO analysis; DOD interviews and documentation. | GAO-15-503

For the other seven case studies included in our review, cost and schedule effects in resolving their disputes were more limited, and in some cases, were not related to operational testing requirements. These disputes also had been, or are expected to be, resolved among DOT&E, the programs, and the operational test agencies. Though not readily quantifiable, for some cases like the Automated Biometrics Identification System and Three-Dimensional Expeditionary Long-Range Radar, we also found indications of benefits coming from resolving their disputes. Several of the programs cited additional costs as a result of DOT&E actions or requirements. Typically, these costs were associated with DOT&E officials requesting changes to test approaches proposed by the military services because they found them insufficient to demonstrate the system's operational effectiveness and suitability. In other cases, program officials stated they experienced cost increases, but they were associated with other factors, such as system development challenges. In the case of the Paladin Integrated Management program, DOT&E's identification of discrepancies with system requirements and operational capability expectations resulted in adjustments to the test approach the Army had sought, which modestly affected test cost. Regarding effects on schedule, officials from three programs—Enhanced Combat Helmet, Ground/Air Task Oriented Radar, and Joint Assault Bridge—stated that potential or realized delays were attributable, in part, to test-related issues. For example, in the Enhanced Combat Helmet program, delays were attributed to test and non-test factors, including changes to the Marine Corps' proposed test approach to conform to recently-established standardized testing protocols and helmet design modifications resulting from manufacturing changes that had degraded helmet performance. The following profiles provide details on the disputes, resolution, and associated impacts for these seven cases.

DOD Automated Biometrics Identification System

To address obsolescence issues with the Automated Biometrics Identification System 1.0—a biometric (e.g., fingerprints) data repository and match capability used to identify potential threats to U.S. military forces and facilities—the Army developed Automated Biometrics Identification System 1.2. However, after several failed attempts to deploy the upgraded system, two military commands requested that the Director, Operational Test and Evaluation (DOT&E) oversee operational test and evaluation for the system. The system’s origins were as a quick reaction capability—not an acquisition program—which posed challenges because no formal operational testing had been planned. In working to establish a plan, the Army’s operational test agency and DOT&E had disagreements about system requirements and initial operational test and evaluation plans. Army and DOT&E officials acknowledged the unique circumstances of the system and pressure to quickly deliver the updated system created substantial tension.

Impact: Due to the lack of upfront operational test planning, no operational test funding was budgeted, so the Army had to reprogram a limited amount of funds to support the testing. DOT&E worked with Army program and test officials to develop a test approach that enabled testing and deployment of the system to meet the program’s needs. Program officials noted that DOT&E was a positive forcing factor in getting the system tested and deployed to meet their schedule needs.

Resolution status: Initial operational testing was completed in 2014 and the Automated Biometrics Identification System 1.2 system has been deployed to users in replacement of the legacy version.

Source: GAO analysis; DOD interviews and documentation. | GAO-15-503

Enhanced Combat Helmet

The Enhanced Combat Helmet program, which responds to an urgent need requirement, had several challenges related to first article test—a process used to determine if the helmet met its contract specifications prior to acceptance. Shortly before the initial first article test, the Director, Operational Test and Evaluation (DOT&E) released a standard test policy for combat helmets that had been developed in coordination with the military services and U.S. Special Operations Command. This new policy, which was in response to past criticism that DOD had received about the testing of personal protective equipment, was a source of frustration for the Marine Corps because it forced changes to test procedures it had intended to use for the Enhanced Combat Helmet. DOT&E noted this policy established minimum standards for testing, ensuring helmets meet a common standard across DOD using an approach employed by commercial manufacturers to balance consumer and producer risk. The helmet failed this first article test, in part because of issues with the test methods used. Using revised test methods, in late 2011 a second first article test demonstrated the helmet met requirements. However, a subsequent manufacturing change degraded helmet performance, leading to the helmet failing small arms testing conducted in June 2012. Once helmet design modifications were made to address shortfalls, a third first article test was completed in April 2013, which the helmet passed. DOT&E's reporting of helmet performance in testing and its potential effect on the health of a wearer was another source of concern for the Marine Corps. Specifically, using input from the Armed Forces Medical Examiner, DOT&E reported that inward deformation of the helmet shell during testing presented a serious risk of injury or death, whereas the Marine Corps stated the implications on health are unknown. The production approach also posed a challenge. The Enhanced Combat Helmet program had intended to expedite production and fielding by having lots comprised of every helmet size. However, the Marine Corps and DOT&E could not identify a viable approach to lot acceptance testing—a quality control test where a sample of helmets is tested from each lot manufactured—that would support lots with mixed helmet sizes.

Impact: Meeting the newly standardized test policy requirements necessitated more funding and time than planned for by the program. Program officials stated they experienced about 12 months in delays and about \$2 million in cost increases during operational testing. Officials stated that meeting DOT&E test protocol requirements contributed to the cost and schedule growth, but helmet performance problems and material changes, contract renegotiations, and test requirements independent from DOT&E oversight were also factors. DOT&E stated that additional risk reduction tests pursued by the program after the initial first article test and delays associated with the helmet developer resolving the manufacturing issues led to the majority of the program delays. DOT&E also indicated that issues with the test procedures were due to the unanticipated behavior of the helmet when shot during testing and emphasized that these issues would also have occurred with the Marine Corps' originally-proposed test procedures. Regarding the health considerations related to Enhanced Combat Helmet test results, DOT&E and the Marine Corps operational test agency recommended additional testing in concert with the medical community to characterize the potential for injury from helmet deformations. For production, the industry best practice of single-size helmet lots has been used, which slowed fielding plans because marine or soldier units are not equipped with the helmets until all needed sizes are available but also reduced the risk of deficient helmets being fielded.

Resolution status: The Marine Corps found the Enhanced Combat Helmet preferable to the existing lightweight helmet and proceeded to full-rate production and fielding.

Source: GAO analysis; DOD interviews and documentation. | GAO-15-503

Ground/Air Task Oriented Radar

At the Under Secretary of Defense for Acquisition, Technology, and Logistics' (USD (AT&L's)) direction, in 2013 the Marine Corps revised the reliability growth program for the Ground/Air Task Oriented Radar—a portable short- to medium-range air defense and air surveillance radar—to address issues discovered during developmental testing. Despite the revisions, the Director, Operational Test and Evaluation (DOT&E) did not approve the test and evaluation master plan (TEMP) ahead of the low-rate initial production decision review, stating that additional changes were needed to make the program's reliability growth program consistent with system requirements, realistic, and achievable by initial operational test and evaluation (IOT&E). The program received low-rate production approval in March 2014, with reliability concerns unresolved and without DOT&E TEMP approval. Shortly thereafter, the Navy commissioned a panel that reviewed the reliability concerns and provided recommendations related to the Ground/Air Task Oriented Radar's reliability requirement and growth plan. In particular, the panel found that selected reliability requirements were disconnected from the program's mission, operational relevance could not be determined, and the reliability growth analyses and predictions were technically deficient. These issues were similar to those identified by DOT&E prior to the low-rate production decision. A material change in semi-conductor technology for the radar system—moving from gallium arsenide to gallium nitride—also created issues between the program and DOT&E. The program planned to complete IOT&E with gallium arsenide radars. However, the preponderance of radars—80 percent—is expected to include the material change. Based on this production approach and because the material change modified the physical characteristics of the radar's aperture and requires software modifications, DOT&E determined that gallium nitride units need to be used to meet its legal requirement to complete IOT&E using production-representative systems. Program officials disagreed with DOT&E's assessment that the material change was significant enough to warrant an IOT&E change, but the test plan was updated to include gallium nitride units in IOT&E. The scope of IOT&E is also being deliberated by the program office, operational test agency, and DOT&E to determine whether the system will be operationally tested in a littoral environment—a primary operating environment for the radar.

Impact: The program office reevaluated system reliability and plans to implement some of the panel recommendations and incorporate them into the TEMP. The radar material change and other programmatic decisions contributed to a delay for IOT&E and full-rate production of about 2 years, and production efficiencies, such as lower unit costs, will not be achieved as planned. Gallium arsenide units are expected to undergo an early fielding test in fiscal year 2017 and IOT&E with gallium nitride units is planned for the following year. Program officials estimated that including a littoral environment in IOT&E could cost about \$20 million. DOT&E noted that test requirements have not yet been established and believes this estimate includes testing components that expand the test scope beyond what is needed.

Resolution status: Program officials are evaluating the reliability requirement for operational relevance and recommendations to improve the reliability growth program. Marine Corps test officials stated that TEMP approval will not be sought until issues related to the reliability program and IOT&E scope are resolved. DOT&E noted the delay in TEMP approval has not affected the program's overall schedule.

Joint Assault Bridge

The Joint Assault Bridge program—which provides an assault bridge-laying capability for armored combat teams—began receiving Director, Operational Test and Evaluation (DOT&E) oversight for live fire test prior to 2011 when it was a Marine Corps program and was placed on operational test oversight in January 2014 when the program had been taken over by the Army. The Army's initial live fire test program sought to gain efficiencies by using data from previous live fire testing of the Abrams tank chassis—a core component of the Joint Assault Bridge system. However, DOT&E stated that changes in number of crew, center of gravity, and other design differences that significantly affect system and crew survivability made data from previous live fire tests insufficient to support an evaluation of Joint Assault Bridge system. Shortly before being added to DOT&E operational test oversight, the Army's operational test agency approved a test plan for the program that included two test events and 40 launch-and-recovery cycles of the system's bridge (e.g., put the bridge down and pick it back up). DOT&E requested further details on the Army's test plan once the program was placed on DOT&E's operational test oversight list, and the Army responded by providing an updated test concept in July 2014 that condensed initial operational test and evaluation (IOT&E) into one event at Aberdeen Proving Ground with 36 launch-and-recovery cycles; the cycle reduction would be mitigated through developmental testing. DOT&E outlined the risks in this test concept, stating that cycles performed in developmental testing would not change the need for at least 65 launch-and-recovery cycles, which DOT&E found through statistical analysis was the minimum number of cycles required to reduce test risk and demonstrate system effectiveness at IOT&E. Additionally, DOT&E, along with the Army's Operational Test Command, found that Aberdeen Proving Ground did not provide an operationally realistic test environment. Fort Hood was recommended as an IOT&E event location.

Impact: The program's live-fire test strategy will include full-up system level, fire survivability, and battle damage assessment and repair tests. The program office estimated a cost of \$7.2 million to reconfigure a development model Joint Assault Bridge for live fire testing. Program officials believed a delay to first unit equipping could occur, with the time needed to incorporate the DOT&E-approved live fire test program and other factors contributing to a potential delay. A minimum of 66 launch-and-recovery cycles are planned for IOT&E at Fort Hood, which reduces risk identified in previous test proposals. Test efficiencies are expected from employing soldiers in IOT&E that are from the potential first unit to be equipped with the Joint Assault Bridge.

Resolution status: The system is scheduled to undergo an integrated test (combined developmental and operational test) at Aberdeen Proving Ground in January 2018 followed by IOT&E at Fort Hood in March 2018.

Source: GAO analysis; DOD interviews and documentation. | GAO-15-503

Paladin Integrated Management

In 2011, the Director, Operational Test and Evaluation (DOT&E) identified discrepancies in the Army's approved requirements document between the stated need for Paladin Integrated Management—a self-propelled howitzer artillery cannon—to operate against specific threats, and protection the system would actually provide based on its technical specifications. In particular, DOT&E stated the system specifications would provide insufficient protection to its crew against existing threats in its intended operational environment. To address this issue, DOT&E recommended an increase in the system's force protection and survivability requirements or a designation that the system would not operate in threat environments, which DOT&E emphasized, were the environments upon which the rationale for acquiring the system were based. In response, the Army updated the requirements and amended its operating concept to eliminate the need to operate in the environment originally intended for the system, and to establish that the system would operate on cleared routes (e.g., routes with minimal improvised explosive device threats). However, DOT&E cited cases from recent operational experience that suggested the Army's updates were inadequate to address threats, and recommended to the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD (AT&L)) that the Army fund, develop, and test an underbody armor kit to address operational threats to the system's underbody, namely improvised explosive devices.

Impact: USD (AT&L) directed the Army to design, develop, and test an underbody armor kit to address operational threats, which if successfully demonstrated, would provide the option to build and deploy those kits to the field as needed. Underbody kit testing will be integrated into the pre-existing developmental, live-fire, and operational test plans and has not significantly altered test costs or timelines. The total number of kits to be procured and fielded will be determined after testing is completed.

Resolution status: The Army stated that the estimated cost of current plans to develop, design, and procure five underbody armor kits is \$1.6 million. Three kits will be used in live fire test and two for initial operational test and evaluation. An evaluation of systems equipped with underbody kits will inform the program's full-rate production decision.

Source: GAO analysis; DOD interviews and documentation. | GAO-15-503

The P-8A Poseidon Multi-Mission Maritime Aircraft

The Navy's P-8A—intended to provide anti-submarine, anti-surface warfare, and intelligence, surveillance and reconnaissance capabilities—is being developed in three increments. The first increment, which is intended to provide unarmed anti-surface warfare, anti-submarine warfare, and intelligence, surveillance, and reconnaissance capabilities, completed initial operational test and evaluation (IOT&E) in 2013 despite the Director, Operational Test and Evaluation (DOT&E) advising the Navy to consider delaying it because of known hardware and software deficiencies. DOT&E officials stated that about a month before IOT&E, the Navy deferred wide-area anti-submarine warfare search testing planned for Increment 1 because it intended to wait until a new system in development was available before putting anti-submarine warfare capability on P-8A. DOT&E acknowledged in its IOT&E report that the Navy had deferred the capability, but reported that the P-8A Increment 1 was unable to execute the full range of anti-submarine warfare mission tasks defined by its original concept of operations, which had not been modified. For anti-submarine warfare search, DOT&E reported that P-8A would only be effective if precise cueing was provided or if a wide-area search capability was integrated into the aircraft. Navy officials took issue with DOT&E holding them accountable for this deferred capability, noting that the deferral was based on the Navy's decision that it did not want to invest in putting the legacy system on the aircraft when the system being developed provided greater capability and was expected to be available in the near future. Increment 2 test planning was another source of disagreement between the program and DOT&E. As P-8A neared a full-rate production decision in 2013, the content of Increment 2 was still in flux, delaying development of its test and evaluation master plan (TEMP). DOT&E did not agree to the Navy's proposals for testing Increment 2, noting the Navy's plans were inadequate for determining operational effectiveness in some operational environments and against the primary threat target. According to DOT&E, 1 month prior to a Defense Advisory Board review, the Navy's operational test agency proposed a new test concept that involved highly-structured, small-area field tests. DOT&E formally notified the Navy that this proposal was also unacceptable due to several technical reasons. DOT&E also informed the Navy that a beyond low-rate initial production report on the IOT&E results for Increment 1 would not be submitted until the Increment 2 TEMP was approved because the wide-area search capability deferred from Increment 1 of the program to Increment 2 is a key reason the Navy is acquiring the P-8A.

Impact: Disagreements over the Increment 2 TEMP did not affect the full-rate production decision. Conducting agreed-to operational testing for Increment 2 will require the Navy to increase planned operational test funding.

Resolution status: The Navy and DOT&E came to an agreement on a test plan for Increment 2 that provides statistical rigor, a sufficient variety of test flights and environments, and end-to-end testing of anti-submarine warfare capability. Once agreement was reached on the Increment 2 TEMP, the Navy was able to proceed to full-rate production for Increment 1 as scheduled.

Source: GAO analysis; DOD interviews and documentation. | GAO-15-503

Three-Dimensional Expeditionary Long-Range Radar

In 2013, the Director, Operational Test and Evaluation (DOT&E) expressed concerns to the Air Force about the operational relevance of, and the ability to test and evaluate, the 720-hour mean-time-between-critical-failure reliability requirement for the Three-Dimensional Expeditionary Long-Range Radar—a long-range, ground-based sensor for detecting, identifying, tracking, and reporting aircraft and missiles. Prior to the decision to start the program’s system development, DOT&E communicated to the Under Secretary of Defense for Acquisition, Technology, and Logistics and the Joint Chiefs of Staff that this requirement was not adequately justified and had a high risk of not being achievable or testable. Based largely on DOT&E’s reliability requirement concerns, senior stakeholders for the program recommended a delay to the system development request for proposal release until the reliability matter was resolved. Air Force officials took issue with what they viewed as DOT&E seeking to change program requirements, particularly because the change was posed shortly before an August 2013 Defense Advisory Board review where the release of a system development request for proposal was expected to be approved. DOT&E stated its intent was to highlight an issue that was likely to cause the program significant problems later on before system development was begun.

Impact: Upon further assessment, the Air Force lowered the reliability requirement to 495 hours, which still allows the system’s availability requirement to be met. Program officials did not indicate any cost impact from the requirement change, but the request for proposal was delayed, as recommended, until the reliability requirement was revised. The lower requirement is more likely to be achievable, and combined with the new strategy, should improve the radar’s performance against its reliability goal by initial operational test and evaluation. Air Force and DOT&E officials stated that the new requirement will likely benefit the long-term performance of the radar.

Resolution status: The program’s system development request for proposal, which included an updated reliability requirement, was released in November 2013.

Source: GAO analysis; DOD interviews and documentation. | GAO-15-503

Agency Comments

We provided a draft of this report to DOD for comment. The department responded that it did not have any formal comments on the report. However, DOD provided technical comments which we incorporated into the report, as appropriate.

We are sending copies of this report to appropriate congressional committees; the Secretary of Defense; the Secretaries of the Army, Navy, and Air Force; and the Director, Operational Test and Evaluation. In addition, the report will be made available at no charge on the GAO Web site at <http://www.gao.gov>.

If you or your staff have any questions concerning this report, please contact me at (202) 512-4841. Contact points for our offices of Congressional Relations and Public Affairs may be found on the last page of this report. Staff members making key contributions to this report are listed in appendix II.

A handwritten signature in black ink, appearing to read 'Michael J. Sullivan', with a stylized, flowing script.

Michael J. Sullivan
Director, Acquisition and Sourcing Management

List of Committees

The Honorable John McCain
Chairman
The Honorable Jack Reed
Ranking Member
Committee on Armed Services
United States Senate

The Honorable Thad Cochran
Chairman
The Honorable Richard J. Durbin
Ranking Member
Subcommittee on Defense
Committee on Appropriations
United States Senate

The Honorable Mac Thornberry
Chairman
The Honorable Adam Smith
Ranking Member
Committee on Armed Services
House of Representatives

The Honorable Rodney Frelinghuysen
Chairman
The Honorable Pete Visclosky
Ranking Member
Subcommittee on Defense
Committee on Appropriations
House of Representatives

Appendix I: Objectives, Scope, and Methodology

The Joint Explanatory Statement to Accompany the National Defense Authorization Act for Fiscal Year 2015 directed GAO to review the oversight activities of the Department of Defense's (DOD) Office of the Director, Operational Test and Evaluation (DOT&E), including how they may affect acquisition programs.¹ Our objectives for this review were to examine (1) the extent to which there have been any significant disputes between DOT&E and acquisition programs over operational testing, and (2) the circumstances and impact of identified operational test-related disputes.

For our work, we reviewed relevant statutes and DOD policies and guidance related to operational testing and DOT&E. To assess the extent to which there have been any significant operational test-related disputes between DOT&E and acquisition programs, as well as the circumstances associated with them, we conducted interviews with acquisition and test officials within the military services—Air Force, Army, Marine Corps, and Navy—and the offices of DOT&E, the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation, and the Joint Chiefs of Staff. We asked officials to provide their perspectives on the extent and circumstances of significant disputes—disputes that they believed were beyond what is typical for programs—related to operational testing that have occurred between acquisition programs and DOT&E. In addition to our interview activities, we formally solicited input and received responses from officials within the military service acquisition executive offices, test and evaluation offices, and operational test agencies that identified programs that had experienced any of the following circumstances since fiscal year 2010:

- significant delays in obtaining DOT&E's approval of test and evaluation master plans (TEMPs);
- significant delays in obtaining DOT&E approval of operational test plans to support initial operational test and evaluation (IOT&E);
- significant disputes over the test assets needed to conduct IOT&E;
- significant disputes with DOT&E related to what requirements were to be tested during IOT&E, such as testing key performance parameters

¹ Joint Explanatory Statement to Accompany the Carl Levin and Howard P. "Buck" McKeon National Defense Authorization Act for Fiscal Year 2015, 160 Cong. Rec. H8671, H8704 (Dec. 4, 2014).

only versus testing to intended mission capabilities in an operational environment;

- significant disagreements over the characterization of IOT&E results that led to a delay in reaching a full-rate production decision;
- significant disputes related to the need to conduct live fire testing or the extent of testing;
- substantially increased costs for operational test completion that were judged unwarranted by programs or the military services; or
- significant disputes over any other elements associated with DOT&E's oversight role, such as decisions by DOT&E to add programs to its oversight list, DOT&E's activities related to operational assessments, or the need to conduct follow-on operational testing and evaluation.

To determine the total number of programs identified as having had significant disputes with DOT&E, we evaluated the collective information gathered from our interviews and formal inquiries against the above criteria. The specific sources that were asked to identify programs with significant disputes resided within the following offices:

- Air Force Operational Test and Evaluation Center;
- Army Test and Evaluation Command;
- Assistant Secretary of the Air Force (Acquisition);
- Assistant Secretary of the Army (Acquisition, Logistics and Technology);
- Assistant Secretary of the Navy (Research, Development and Acquisition);
- Commander, Operational Test and Evaluation Force (Navy);
- Deputy Under Secretary of the Army (Test and Evaluation);
- Deputy Assistant Secretary of the Navy (Test and Evaluation);
- Director, Air Force Test and Evaluation;
- Marine Corps Operational Test Activity; and
- Marine Corps Systems Command.

We combined the input received from these sources to form a list of all programs reported as having experienced significant disputes with DOT&E. We then reviewed the information gathered on the disputes for each program to verify that one or more of the stated criterion was

associated with each program by at least one source. For the purposes of our review and reporting, all programs that met this standard—42 in total—were considered to have had a significant dispute with DOT&E related to operational testing.

In addition to collecting information on programs with significant disputes, we used DOT&E's annual reports from fiscal years 2010 through 2014 to compile a complete list of programs—454 in total—that were on the DOT&E oversight list during that time frame.

To obtain a better understanding of the circumstances and impact of operational test disputes identified by the military services, we elected to conduct a case study analysis of a select number of the programs. Based on an assessment of the information we collected on the 42 programs identified with significant disputes and discussions with military service and DOT&E officials, we judgmentally selected 10 cases for in-depth review and analysis. The 10 cases we selected were considered to be among the most significant disputes that occurred in each of the military services in recent years. The cases selected include:

- CVN 78 *U.S.S. Gerald R. Ford* Class Aircraft Carrier, Navy
- DOD Automated Biometrics Identification System, Army
- DDG-51 Flight III Destroyer, AN/SPY-6 Radar, Aegis Modernization and Self-Defense Test Ship, Navy
- Enhanced Combat Helmet, Marine Corps
- F-35 Joint Strike Fighter and the Electronic Warfare Infrastructure Improvement Program, DOD
- Ground / Air Task Oriented Radar, Marine Corps
- Joint Assault Bridge, Army
- P-8A Poseidon Multi-Mission Maritime Aircraft, Navy
- Paladin Integrated Management, Army
- Three-Dimensional Expeditionary Long-Range Radar, Air Force

For each case, we interviewed officials from program offices, program executive offices, or both; operational test agencies; and the office of DOT&E. We also reviewed programmatic documentation, such as test and evaluation master plans, operational test plans, program briefs, memoranda, and operational test reports, as well as other information documenting program cost, schedule, and performance. We assessed

the information obtained from program and test officials in order to determine more fully the circumstances of each dispute and any corresponding impact to the programs or DOD in general. Our case study programs are representative of the types of disputes identified overall by military service officials. However, our case study findings are not generalizable to the total population of disputes communicated to us or to other defense acquisition programs.

We conducted this performance audit from July 2014 to June 2015 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.

Appendix II: GAO Contact and Staff Acknowledgments

GAO Contact

Michael J. Sullivan, (202) 512-4841 or sullivanm@gao.gov

Staff Acknowledgments

Principal contributors to this report were John Oppenheim, Assistant Director; Laura Greifner; Seth Malaguerra; Sean Merrill; Andrew Redd; and James Tallon. Other key contributors included Nathan Foster, Victoria Klepacz, and Sylvia Schatz.

Appendix III: Accessible Data

Data Table for Figure 1: Programs on the Director, Operational Test and Evaluation (DOT&E) Oversight List, Fiscal Years 2005-2014

Number of programs

Fiscal year	Operational test oversight list	Live fire oversight list
2005	279	96
2006	293	106
2007	288	108
2008	322	122
2009	322	128
2010	348	132
2011	311	118
2012	327	127
2013	312	117
2014	309	121

Source: GAO analysis of DOT&E annual reports for fiscal years 2005-2014. | GAO-15-503

Note: Programs can be on DOT&E oversight for operational testing, or live fire testing, or both.

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 GAO's website (<http://www.gao.gov>). Each weekday afternoon, GAO posts on its website newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to <http://www.gao.gov> and select "E-mail Updates."

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, <http://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 [Facebook](#), [Flickr](#), [Twitter](#), and [YouTube](#). Subscribe to our [RSS Feeds](#) or [E-mail Updates](#). Listen to our [Podcasts](#). Visit GAO on the web at www.gao.gov.

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Website: <http://www.gao.gov/fraudnet/fraudnet.htm>

E-mail: fraudnet@gao.gov

Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Katherine Siggerud, Managing Director, siggerudk@gao.gov, (202) 512-4400, U.S. Government Accountability Office, 441 G Street NW, Room 7125, Washington, DC 20548

Public Affairs

Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800, U.S. Government Accountability Office, 441 G Street NW, Room 7149, Washington, DC 20548