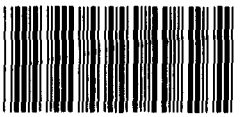
United States General Accounting Office 138560

GAO

Testimony



For Release
on Delivery
Expected at
1:30 p.m.
Thursday,
May 4, 1989

DOD Test and Evaluation

Statement of
Eleanor Chelimsky
Assistant Comptroller General
Program Evaluation and
Methodology Division

Before the
Military Reform Caucus

Members of the Military Reform Caucus:

It is a pleasure to be here today to participate in this panel on the status of testing and evaluation in the Department of Defense (DOD).

In GAO's Program Evaluation and Methodology Division, we've been looking at both developmental and operational test and evaluation for the last 7 years, along with live-fire testing. As we stated in our 1987 report, we fully support the need for full-up live-fire testing--that is, the testing of complete systems with combustibles on board. Full-up testing is the only method providing direct observation of damage under realistic conditions, giving it a unique, important advantage over other methods. We reported then that lack of targets had been a problem, that the technical capability to do full-up live-fire testing was not yet well enough developed, and that disputes between proponents of full-up testing and advocates of computer modeling were slowing progress. We know the live-fire testing office has been seriously trying to come to grips with all of these issues.

In our 1983 report on DOD's joint operational test and evaluation, we found that unrealistic test conditions, together with problems of analysis and reporting, raised serious questions about the validity of the evaluations conducted jointly by the services. We reported at that time that at least some of these problems may have been due to the organizational placement of the

test program under the Director for Defense Research and Engineering, and we were hopeful that the newly enacted legislation naming an independent Director for Operational Test and Evaluation would reduce the serious quality problems we had found.

As you know, the Congress established the Office of the Director for Operational Test and Evaluation (DOT&E) to effect several reforms concerning operational testing. Prominent among the reform objectives were: independent oversight and coordination of the military services' planning and execution of operational tests; independent evaluation of the results of those tests; and objective reporting of the results to decisionmakers in DOD and the Congress. A fundamental concern was that weapons were not being tested thoroughly or realistically and that complete and accurate information about them was not being disseminated.

In part to determine if in fact things had improved since the establishment of DOT&E, four members of the House of Representatives asked us in 1987 to review the program. This culminated in our July 1988 report, Weapons Testing: Quality of DOD Operational Testing and Reporting. Our study addressed two evaluation questions: (1) What is the methodological adequacy of operational test and evaluation under DOT&E oversight?, and (2) what is the quality of DOT&E dissemination of information to the Congress?

To address these questions, we reviewed relevant documentation on the operational test and evaluation of six major, conventional weapon systems which had reached, or were scheduled to reach, the full production milestone by the end of FY 1987, as well as congressional testimony, DOD regulations, and outside literature on the conduct and reporting of test and evaluation in general.[1] We also interviewed DOD officials and outside experts in operational testing. We developed a standardized assessment framework to evaluate each system, after which we synthesized the information across systems to yield overall findings and conclusions. The results are generalizable to the universe of major, conventional weapon systems that reached the B-LRIP milestone by the end of FY 1987. They are not generalizable to strategic systems or to systems which had not yet reached the full production milestone in that time frame.

With regard to the methodological adequacy of operational test and evaluation under DOT&E oversight, we found significant problems and limitations in the planning, execution, realism, analysis, and reporting by the service test agencies for the six systems we reviewed (see table 1 for a listing of problems in test realism alone). Some of these problems and limitations were unavoidable due to time, resource, or safety constraints; numerous others were not. Our conclusion was that for major,

---

[1] The six systems were systematically selected from a universe of 10 eligible systems; the specific selection criteria are described in the report.

conventional systems that reached the full production milestone by end-FY 1987, the operational test and evaluation being conducted under DOT&E oversight was not methodologically adequate to assess the effectiveness and suitability of those systems. Instead, the findings have tended to show more favorable assessments than are likely to be found when the weapons are employed in combat. The danger here is that this can lead to the funding of weapon systems whose operational effectiveness and suitability have not been demonstrated.

With regard to the quality of DOT&E dissemination of information to the Congress, each of the official DOT&E reports that we reviewed contained incomplete or inaccurate statements, and most contained both (see table 2). In addition, the majority of DOT&E's favorable overall assessments of testing adequacy and of system effectiveness and suitability were not supported by the evidence. The omissions, inaccuracies, and overall assessments consistently presented a more favorable presentation to the Congress of test adequacy and system performance than was warranted by the facts. We concluded, therefore, that for major, conventional systems that reached the full production milestone by end-FY 1987, DOT&E's dissemination of information to the Congress has not provided a complete and accurate picture of weapons performance.

As I noted earlier, there are some problems and limitations in operational test and evaluation that cannot be avoided, along with many that can be. But even if they were all unavoidable, there is no reason why they should not be reported completely and accurately. Knowing all the limitations to the test and evaluation findings is critical to the Congress in making weapon funding decisions.

The overall conclusion of our report was, therefore, that both the conduct and reporting of operational test and evaluation under DOT&E oversight had fallen short of the objectives sought by the Congress when it established the office.

Our July 1988 report on DOT&E offered no recommendations for changing the current law because we think the 1983 legislation is adequate when combined with DOD's own directives. Together, we believe the two provide the necessary organizational structure and guidance for the conduct and reporting of sound operational test and evaluation. The real problem that needs to be addressed, in our view, is twofold: first, how to ensure that testing is the best it can be: for example, seeing to it that methodological biases are either removed or controlled for, that available resources are used, and that assessments of weapon performance are not overly favorable. Second, how to ensure that the best use is made of test and evaluation findings: for example, improving the completeness and accuracy of the information DOT&E disseminates to

the Congress and thus allowing budget decisions to be properly supported by real knowledge of weapon system effectiveness and its limitations.

Table 1: Significant Problems and Limitations in Test Realism

	Army systems		Navy systems		Air Force systems	
Assessment questions	AHIP	Aquila	TLAM/C	DDG-51	IR Maverick	LANTIRN
Realism						
Operated by typical operational units?		X				X
Operated by typical operational personnel?	X	X	X		X	
Supported by typical support units?	X	X		X	b	X
Supported by typical support personnel?	X	X		X	b	X
Equipment put under realistic stress?	X	X	X	X	X	X
Personnel put under realistic stress?	X	X		X	X	X
Realistic combat tactics employed?	X	X	X	X	X	X
Physical environment approximates intended ranges?	X	X	X		X	X
Target systems approximate actual targets, realistically employed?	X			X	X	X
Threat systems approximate actual threat, realistically employed?	X	X	X	X	X	a
Tested system production representative and prepared for test in a realistic manner?				X		X

Note: empty cells signify "no significant problems or limitations found." X signifies "one or more significant problems or limitations found." a signifies "insufficient information to evaluate." b signifies "not applicable."

Table 2:
Significant Problems in Completeness & Accuracy of DOT&E Reporting

	Army systems		Navy systems		Air Force systems	
DOT&E reporting	AHIP	Aquila	TLAM/C	DDG-51	IR Maverick	LANTIRN
OT&E adequacy						
Statements complete?	X	a	X	X	X	X
Statements accurate?	X	X		X	X	X
System effectiveness and suitability						
Statements complete?		a	X	X	X	X
Statements accurate?	X	a	X	X		X

Note: empty cells signify "no significant problems or limitations found." X signifies "one or more significant problems or limitations found." a signifies "insufficient information to evaluate."