Operational Testing and the Mythical Commercial-off-the-Shelf Aircraft: The Tale of the T-3A Firefly

Christopher L. Harlow

Follow this and additional works at: https://scholar.smu.edu/jalc

Recommended Citation
https://scholar.smu.edu/jalc/vol75/iss3/1

This Article is brought to you for free and open access by the Law Journals at SMU Scholar. It has been accepted for inclusion in Journal of Air Law and Commerce by an authorized administrator of SMU Scholar. For more information, please visit http://digitalrepository.smu.edu.

CHRISTOPHER L. HARLOW*

I. INTRODUCTION

IN 1995, THE UNITED STATES Air Force purchased the T-3A Firefly (T-3) from Slingsby Aviation Limited. The Air Force intended this commercial-off-the-shelf (COTS) trainer aircraft to be used in its flight screening program. The Air Force hypothesized that earlier candidate exposure to aerobatic flight would reduce pilot eliminations during the later, more expensive, phases of training. As a COTS acquisition, the T-3 was not subjected to traditional operational test and evaluation (OT&E). Additionally, because the acquisition involved expenditures of only $40 million, the program fell well below the statutory financial thresholds mandating OT&E.

The failure to operationally test the T-3 produced disastrous results for the Air Force. Within two years, six young pilots were dead, and the Air Force was forced to remove the T-3 from its inventory. Following a failed attempt to sell its fleet of T-3s for parts, the Air Force ultimately disposed of the planes as scrap metal. A variety of factors, ranging from personnel limitations to unique geographic issues, were blamed for the T-3’s failure. Ultimately, the exact cause of the T-3’s failure is irrelevant because full and early OT&E could have exposed each of these shortcomings.

Two basic mechanisms protect American soldiers from the negative consequences of ill-suited and ineffective government

* Christopher Harlow is a litigation associate in the D.C. office of Sonnenschein Nath & Rosenthal. Prior to attending The George Washington University Law School, he served as an Air Force captain and test manager at the Air Force Operational Test & Evaluation Center. Harlow is a 1998 graduate of the United States Air Force Academy. In 1997, he was assigned to the 557th Flying Training Squadron where he flew the T-3A Firefly.
purchases. The first, OT&E, is generally required prior to broad production of "major defense acquisitions" and combat-designated programs. This testing is conducted by service-specific teams with both acquisition and technical expertise. The inclusion of potential operators in pre-fielding testing provides an important layer of protection for all future soldier-operators.

Despite its benefits, OT&E is not required for relatively inexpensive noncombat COTS products. These products introduce the second procurement mechanism protecting American soldiers—contractor liability. With respect to these items, the developer remains financially liable for injuries caused by COTS products the government has not substantially altered. Contractor liability and OT&E combine to protect American soldiers from ineffective and unsuitable products in most situations. When the government-contractor defense precludes liability, OT&E ensures products are "effective" and "suitable." Conversely, OT&E is generally not required for COTS products subject to contractor liability, presumably because the contractor's financial incentives will produce effective and suitable products.

While most COTS purchases are relatively mundane, the Air Force has purchased noncombat COTS products whose soldier-operators are not adequately protected by either mandatory OT&E or contractor liability. In the case of the T-3, this gap cost the Air Force nearly $40 million and six young pilots their lives. A slight modification in the statutory language of 10 U.S.C. § 2399 would fill this gap by expanding the scope of mandated OT&E. This conclusion is supported by analyses of the purpose and capabilities of OT&E, as well as the difficulties typically encountered with COTS aircraft. The tale of the T-3 provides a poignant illustration of how expanded OT&E could effectively address the inherent problems of COTS aircraft. The broader application of mandatory OT&E could prevent similar tragedies in the future.

II. OPERATIONAL TESTING

A. OPERATIONAL TESTING STATUTORY BACKGROUND

While the Department of Defense (DOD) requires OT&E of particularly expensive acquisitions, the statutory language omits potentially perilous items that do not satisfy the financial thresholds.\(^1\) The DOD mandates OT&E of major defense acquisitions

prior to low-rate initial production. Programs with total procurement expenditures exceeding $540 million or research, development, and test expenditures exceeding $115 million are defined as "major" acquisitions for operational testing purposes. These expenditures include all "elements that will function together to produce the capabilities required to fulfill a mission need . . . including hardware, equipment, [and] software." In mandating OT&E for acquisitions exceeding these financial thresholds, Title X assumes a broad systems perspective. The importance of this definition and the inadequacy of Title X's OT&E mandate have been revealed by the Air Force's COTS aircraft failures, particularly the T-3.

In addition to definitive financial thresholds, agency heads are authorized to designate an acquisition as major and thus require operational testing. Despite this authority, the substantial costs of OT&E effectively exempt most programs below the financial thresholds. This situation is illustrated by the tragic example of the T-3. With expenditures of only $40 million, the T-3 was exempt from most OT&E requirements despite the inherent dangers associated with trainer aircraft.

Notwithstanding the financial threshold requirements, the Secretary of Defense may waive OT&E for certain aircraft. The Secretary is authorized to designate certain acquisitions as "pilot programs" exempt from OT&E. Included within this exemption are "commercial-derivative aircraft," defined as "any aircraft (including spare parts, . . .) that is or was of a type customarily used in the course of normal business operations for other than Federal Government purposes." These aircraft must be type certified by the Federal Aviation Administration (FAA) and must have been sold or leased in the commercial marketplace. These regulations establish a broad OT&E exemption for aircraft deemed to be COTS products. In combination with the

---

2 § 2399(a)(1).
3 § 2302d(a)(1)-(2).
4 § 2302(5).
5 See infra Part IV.C. (discussing the T-3's failures).
6 §§ 2399(a)(2)(A), 2302(5).
7 See infra Part IV.C.
8 See infra note 107 and accompanying text.
10 Id.
12 Id.
statutory financial thresholds, these regulations serve to exempt most COTS acquisitions from OT&E—including trainer aircraft like the T-3. These exemptions prevent exposure of many of the problems inherent to COTS aircraft—problems OT&E is ideally suited to reveal.

**B. FUNDAMENTAL OPERATIONAL TESTING PRINCIPALS**

Operational testing provides a critical safeguard for acquisitions exceeding the statutory financial thresholds or designated as major by the responsible agency. Operational testing ensures that systems satisfy the critical objectives established by future operators.\(^{13}\) This entails an assessment of the “effectiveness” and “suitability” of a potential acquisition “under realistic operational conditions.”\(^{14}\) The “effective” and “suitable” standards provide operational test agencies with a two-pronged framework for evaluating a system under acquisition consideration.

In general terms, OT&E assesses the ability of the system and its future operators to successfully execute and sustain a particular mission in an operational environment.\(^{15}\) Effectiveness directly addresses the system’s ability to perform its assigned mission.\(^{16}\) Suitability evaluates the ability of future operators to perform this mission with the system, as well as the system’s ability to integrate into the existing infrastructure and doctrine.\(^{17}\) Suitability evaluation is enabled by employing typical military users during operational testing.\(^{18}\) Combining a realistic mission environment with typical future operators allows OT&E to identify a wide range of operational issues prior to system fielding.

The Air Force primarily assigns these tasks to the Air ForceOperational Test and Evaluation Center (AFOTEC).\(^{19}\) For programs requiring OT&E, AFOTEC is involved at the earliest de-

---

\(^{13}\) U.S. DEp’T OF DeF., DEPARTMENT OF DEFENSE INSTRUCTION 5000.2-R: MANDATORY PROCEDURES FOR MAJOR DEFENSE ACQUISITION PROGRAMS (MDAPS) AND MAJOR AUTOMATED INFORMATION SYSTEMS (MAIS) ACQUISITION PROGRAMS ¶ C3.6.1 (2001).

\(^{14}\) Id.

\(^{15}\) BERNARD FOx ET AL., RAND PROJECT AIR FORCE, TEST AND EVALUATION TRENDS AND COSTS FOR AIRCRAFT AND GUIDED WEAPONS 18 (2004).

\(^{16}\) Id.

\(^{17}\) Id.


\(^{19}\) FOX ET AL., supra note 15, at 18.
velopment stages.\textsuperscript{20} Through early involvement and frequent, iterative testing, AFOTEC provides critical feedback to both developers and operators.\textsuperscript{21} Air Force policy defines the purpose of OT&E as providing a mechanism "to mature system designs, manage risks, identify and help resolve deficiencies as early as possible, and ensure systems are operationally mission capable (i.e., effective and suitable)."\textsuperscript{22} Organizing OT&E as "a series of well-defined gates" enables AFOTEC to minimize the government's acquisition risk.\textsuperscript{23} For major acquisitions, AFOTEC provides the Air Force with a critical safeguard against defective products.\textsuperscript{24}

While OT&E is an effective means of exposing system deficiencies, it also provides critical evaluation of the system's interaction with operators and the service in general. Air Force instructions describe this additional evaluation as providing "information on organization, personnel requirements, doctrine, and tactics."\textsuperscript{25} This broad perspective of OT&E derives from the statutory definition of "major," encompassing all elements "required to fulfill a mission need."\textsuperscript{26} OT&E frequently reveals performance-specific deficiencies, as well as deficiencies related to personnel training, doctrine, and organization.\textsuperscript{27} Early exposure of these problems spurs system improvement and ultimately protects military operators from ineffective and unsuitable products.

The widespread use of COTS acquisitions creates significant problems for traditional OT&E. With the advent of COTS acquisitions, the testing community recognized the unique challenges of testing these products.\textsuperscript{28} Despite the commercial prevalence of these products, the testing community realized that OT&E could not be eliminated "simply because these items

\textsuperscript{20} Id.
\textsuperscript{21} Id.
\textsuperscript{22} U.S. AIR FORCE, AIR FORCE INSTRUCTION 99-103: TEST AND EVALUATION 6 (2004).
\textsuperscript{23} Gregory L. Barnette, Test and Evaluation in a Dynamic Acquisition Environment, DEF. ACQUISITION REV. J., 337, 339 (2004); see also id. at 337–48.
\textsuperscript{24} Id.
\textsuperscript{27} See CHRISTOPHER L. HARLOW & JOHN D. PIEGZIK, DCAPES OPERATIONAL ASSESSMENT FINAL REPORT (2001) (demonstrating that operational testing exposes these types of deficiencies).
came from pre-established sources and some pre-existing data is available." Consequently, the test community insisted that military-unique applications of COTS products be tested prior to fielding.30

Despite these concerns, the test community conceded that in some situations existing contractor data could be used to reduce the scope and expense of government testing.31 AFOTEC policy specifically recognizes this conflict by requiring only sufficient testing of COTS products focused on unique military applications.32 Despite the importance of testing COTS products, the statutory requirements combined with political pressures often reduce the likelihood of OT&E.33

III. OPERATIONAL TESTING OF COMMERCIAL-OFF-THE-SHELF PRODUCTS

Many of the benefits derived from COTS acquisitions directly conflict with operational testing. In cases of potentially perilous acquisitions, like the T-3, this conflict exposes military operators to fatal program deficiencies. COTS acquisitions frequently encounter integration, functional, and data availability issues.34 These problems have had particularly pronounced consequences for COTS aircraft.35 While these issues present significant obstacles to COTS acquisitions, OT&E could identify these problems earlier and effectively prevent potentially lethal results.

A. COMMERCIAL-OFF-THE-SHELF ACQUISITIONS BY THE DEPARTMENT OF DEFENSE

A COTS product is generally defined as one "customarily used for non-governmental purposes" by the general public.36 Additionally, publicly used products requiring only minor modifications are considered commercial items.37 Modifications are

29 Id.
30 Id. at 38.
31 Id.
35 Id.
36 C. Albert & E. Morris, Office of the Sec'y of Def., Commercial Item Acquisition: Considerations and Lessons Learned 3 (2000).
37 Id.
limited to those not significantly altering the purpose or "essential physical characteristics" of the item. These definitions enable the government to acquire and use products, or their closely related brethren, to fulfill governmental needs. The acquisition of COTS products typically reduces production time, expedites new technology use, lowers life-cycle costs, and creates a wider support base within the commercial sector. The benefits of COTS acquisitions, however, have been tempered by several generally applicable integration issues.

1. Product Integration Issues

Product integration creates several unique problems for the DOD. Rarely does a single contractor offer a consolidated commercial product individually capable of satisfying a requirement. As a result, a DOD program typically integrates items from multiple vendors within a custom-built system. Combining items from various sources ultimately impacts the performance of the completed system, and many DOD programs have failed for lack of careful consideration of integration issues. The prevailing use of COTS products typically requires integration of several complex products, and their unpredictable interactions often doom DOD acquisitions.

2. Functional Disparities

The functional differences between governmental and public use present another unique obstacle to COTS acquisitions. Successful public use of an item may involve functionality that is unimportant to the government, while omitting other critical functions. Frequently, a gap exists between the intended government usage and the prevailing commercial use. When this functional gap is large, COTS acquisition may be inappropriate. In many situations, this gap cannot be remedied by a minor modification. Similar to product integration issues, the

---

38 Id. at 24.
39 Id. at 2.
40 See generally id. at 3-5.
41 Id. at 4.
42 Id. at 6.
43 Id.
44 Id. at 5.
45 Id. at 7.
46 Id. at 8.
47 Id. at 8 n.17.
functional gap encountered in DOD product usage often inhibits the application of COTS products to military environments.48

3. Data Availability Issues

Limitations in data provided by the vendor present another problem for COTS acquisitions. While evaluating their own products, vendors may not produce detailed test data, and often this information “exists only in the minds of [their] engineers.”49 Though this practice may not rise to the level of concealment, the lack of information certainly hampers subsequent government evaluation.50 The lack of adequate evaluation data, combined with functional shortcomings and integration issues, hinders the acquisition of effective COTS products. All three of these shortcomings were readily apparent in the Air Force’s doomed acquisition of the T-3A Firefly and introduce unique operational testing issues.51

B. Operational Testing Issues for Commercial-off-the-Shelf-Products

Political and economic pressures have led DOD leaders to streamline the acquisition process and increase consideration of COTS products. In many acquisitions, OT&E represents a significant portion of program expenditures.52 Specifically, OT&E accounts for over 20% of total developmental costs for fixed wing aircraft.53 Recent OT&E expenditures for these aircraft range from 17% for the F-14 to 65% for the F-22.54 These expenditures have led many to question the role of OT&E in COTS acquisitions.

Consideration of COTS products has increased, in part, due to the savings realized from decreased testing.55 Of course, this benefit is reduced by the problems inherent to COTS products discussed earlier.56 Despite these shortcomings, the perception

48 Id.
49 Id. at 15.
50 Id. at 12.
51 See infra Part IV.C. (discussing the shortcomings of the T-3A Firefly).
52 Fox et al., supra note 15, at xv n.1.
53 Id.
54 Id. at 94.
55 Id. at 36.
56 See supra Part III.A.
prevails that COTS acquisitions reduce testing requirements.\textsuperscript{57} This view is reinforced by the fact that “[O]T&E expenditures generally occur in the later stages of development when cost overruns and schedule slips have already occurred.”\textsuperscript{58} These factors directly led to the subordination of OT&E to the perceived benefits of COTS acquisitions.

Department of Defense policies reflect this changing landscape. DOD instructions mandate the use of commercial items “to the maximum extent possible” because many commercial items “have application[s] to DOD systems.”\textsuperscript{59} While complete COTS aircraft remain rare, the use of commercial aircraft components is increasingly common.\textsuperscript{60} These components are deemed less expensive than custom military items because they shift developmental costs and technological risks to the private sector.\textsuperscript{61} However, subsequent studies have revealed that these products often require military-specific testing.\textsuperscript{62} The complex nature of governmental acquisitions negates many of these perceived benefits and complicates the acquisition process.

C. COMMON MISCONCEPTIONS OF COMMERCIAL-OFF-THE-SHELF BENEFITS

Independent studies conducted by NASA and the RAND Corporation identified several common misperceptions concerning COTS acquisitions for complex government programs.\textsuperscript{63} A 2001 NASA study revealed several COTS misconceptions.\textsuperscript{64} These include the following assumptions: COTS modifications can be executed sooner than new products can be developed; appropriate COTS products are readily available; and literature and data provided by the COTS vendor is accurate and applicable.\textsuperscript{65} Reconsidering these assumptions led some at NASA to conclude that in many situations COTS solutions are not necessarily less risky or less expensive than custom development.\textsuperscript{66}

\textsuperscript{57} Paul D. Gutierrez, Commercial or Non-Developmental Item Acquisition Strategy: A Look at Benefits vs. Risks—Test and Evaluation, 31 PROGRAM MANAGER 1, 1–2 (May 2002).

\textsuperscript{58} Fox et al., supra note 15, at 1.

\textsuperscript{59} U.S. Dep’t of Def., supra note 13, ¶ C2.9.1.4.2.1.

\textsuperscript{60} Fox et al., supra note 15, at 36.

\textsuperscript{61} Id.

\textsuperscript{62} See generally Sec’y of the Air Force, supra note 28.

\textsuperscript{63} Gutierrez, supra note 57, at 3.

\textsuperscript{64} Id.; see also Fox et al., supra note 15.

\textsuperscript{65} Gutierrez, supra note 57, at 3.

\textsuperscript{66} Id.
These conclusions parallel more recent findings in a similar study by the RAND Corporation concerning COTS aircraft.\textsuperscript{67} This study built upon the NASA findings and provides valuable insight into the problems encountered by the Air Force in its acquisition of the T-3.\textsuperscript{68} Specifically, in 2004, the RAND Corporation conducted a study on behalf of the Air Force to investigate the common belief that COTS acquisitions reduce test expenditures.\textsuperscript{69} The RAND study was commissioned, in part, due to the increased reliance on COTS aircraft and components.\textsuperscript{70} This study revealed that profound differences in military and commercial operations negated many of the perceived benefits of COTS acquisitions.\textsuperscript{71} The RAND study found that COTS aircraft and aircraft components often required similar levels of testing because of these differences.\textsuperscript{72}

Furthermore, the RAND study revealed that for many programs, COTS acquisition strategies actually increased testing because DOD requirements spurred major redesigns.\textsuperscript{73} System redesigns present twin risks to military acquisition programs. First, program managers may not realize that additional testing is required to evaluate the interaction of added technology with the existing system.\textsuperscript{74} These system redesigns often had unanticipated collateral effects on the existing product.\textsuperscript{75} If additional testing is not performed, military operators may be given ineffective or unsuitable products. In the case of aircraft and aircraft components, these flaws could be fatal.\textsuperscript{76} Second, if additional testing is conducted, expenditures increase and essentially negate the financial benefits of COTS acquisitions.\textsuperscript{77}

The RAND study concluded that "[e]ven well-proven commercial products" must be tested in the unique military environment.\textsuperscript{78} While this added testing reduces the financial benefits of COTS acquisitions, it reduces the risk that military personnel will operate ineffective, unsuitable, and potentially lethal prod-

\begin{itemize}
\item \textsuperscript{67} See generally Fox et al., supra note 15.
\item \textsuperscript{68} Id.
\item \textsuperscript{69} Id. at 4.
\item \textsuperscript{70} Id. at xvi.
\item \textsuperscript{71} Id. at 36.
\item \textsuperscript{72} Id.
\item \textsuperscript{73} Id.
\item \textsuperscript{74} Gutierrez, supra note 57, at 3.
\item \textsuperscript{75} Albert & Morris, supra note 36, at 20.
\item \textsuperscript{76} See infra Part IV.C (discussing the fatalities associated with the T-3).
\item \textsuperscript{77} Gutierrez, supra note 57, at 1-2.
\item \textsuperscript{78} Fox et al., supra note 15, at 39.
\end{itemize}
ucts. The 2001 NASA study provided an initial framework to question the basic underlying assumptions of COTS acquisitions. The 2004 RAND study specifically applied many of these questions to COTS aircraft. The story behind the Air Force’s purchase of the T-3 provides a salient illustration of how these problems can manifest in a specific acquisition with lethal results.

IV. THE FAILURE OF THE SLINGSBY T-3A FIREFLY

The acquisition of the T-3 provides an illustrative case study of the problems inherent to COTS aircraft and the importance of OT&E. With respect to the T-3, senior leadership pressure and time constraints combined to virtually eliminate OT&E entirely. Similar to the problems discussed in the RAND study, the failure to operationally test the T-3 left several significant problems unexposed. Ultimately, this failure produced disastrous results.

A. T-3A ACQUISITION STRATEGY AND BACKGROUND

1. T-3 Background

In an attempt to improve pilot training, the Air Force replaced its existing flight screener with the T-3A Firefly in 1995. In 1952, the Air Force instituted a centralized flight screening program designed to reduce student elimination during later phases of pilot training. The early flight exposure of Air Force Academy cadets and recent ROTC graduates enabled the service to identify candidates incapable of successfully completing pilot training. This early elimination allowed the Air Force to forego the more expensive, later phases of pilot training and redirect these officers to other career fields.

79 Gutierrez, supra note 57, at 3.
80 See generally Fox et al., supra note 15.
82 Id.
84 Id. at 6.
86 Id.
The Air Force operated its flight screening program for forty-two years without noteworthy difficulties.\(^{87}\) The Air Force Academy cadets began flying the T-41, the T-3's predecessor, in 1964.\(^{88}\) Despite the inherent inability and inexperience of thousands of cadets, the T-41 was flown over the Academy Airfield for thirty years without a single fatal accident.\(^{89}\) Despite its spotless safety record, the T-41 was viewed by Air Force leaders as an overly conservative plane incapable of effectively screening pilot candidates.\(^{90}\)

In the 1980s, a movement to replace the T-41 took hold within the Air Force.\(^{91}\) The leading proponent, former Air Force Chief of Staff General Merril McPeak, expressed his disdain for the T-41: "The T-41 is your grandmother's airplane. Our mission is to train warrior-pilots, not dentists to fly their families to Acapulco."\(^{92}\) According to General McPeak, the T-41 did not effectively "pinpoint those cadets who have the basic aptitude to become Air Force pilots."\(^{93}\)

Under the leadership of Lieutenant General Oaks, the Air Force undertook a general review of the T-41 that ultimately recommended its replacement.\(^{94}\) A 1989 Broad Area Review concluded that the T-41 was an ineffective screener because of "its high-wing design and inherent limitations."\(^{95}\) Adhering to the principles espoused by General McPeak, the report concluded that a fully "aerobatic trainer" was needed to screen candidates for the rigors of follow-on pilot training.\(^{96}\) Flight screening would no longer merely evaluate a candidate's ability to understand general flight principles.\(^{97}\) By exposing cadets to earlier, more rigorous flight screening, the T-41's successor would produce more skilled Air Force pilots and further reduce late-training eliminations.

\(^{87}\) Thompson, supra note 81.
\(^{88}\) Id.
\(^{89}\) Baker, supra note 85, at 53–57.
\(^{90}\) Thompson, supra note 81.
\(^{91}\) Id.
\(^{92}\) Id. at 6.
\(^{93}\) Id.
\(^{94}\) SEC'Y OF THE AIR FORCE INSPECTOR GEN., supra note 83, at 3.
\(^{95}\) Id.
\(^{96}\) Id.
\(^{97}\) Baker, supra note 85, at 53.
2. **T-3 Acquisition Strategy**

The Air Force considered several acquisition avenues, ultimately settling on a commercial aircraft sold by the British company Slingsby Aviation Limited. The initial Statement of Need adhered to General McPeak's desire for a sexier flight screener by requiring the plane to allow evaluation of a candidate's ability to react quickly and accurately while flying complex maneuvers. Though leasing an aircraft was initially considered, the Air Force ultimately decided to purchase a flight screener from a commercial source. This decision stemmed from the availability of several commercial aerobatic trainers and DOD policy encouraging COTS acquisitions. In support of its decision, the Air Force asserted that a COTS screener would reduce expenditures because it would require less extensive OT&E. The common misperception that a COTS aircraft could be inserted into a military environment without extensive testing fueled the decision to buy an aerobatic COTS trainer. This decision and the subsequent OT&E exemption resulted in tragedy for the Air Force and the youngest of its pilots.

In September 1991, the Air Force released a Request for Proposal. Seven months later, Slingsby Aviation Limited was awarded the flight screener contract for its aerobatic COTS aircraft—later designated the T-3A Firefly. Slingsby contracted to sell the Air Force the T-3 at a unit price of $295,000, with total expenditures approaching $40 million. The total estimated price reflected the Air Force's belief that the T-3 would require less OT&E because it was Type Certified by the FAA, and it maintained Standard Airworthiness Certificates. This belief coincided with the statutory and regulatory positions of the DOD that commercial aircraft certified by the FAA require re-
duced testing. The T-3 presented the Air Force with an opportunity to test its assumption that COTS aircraft, certified by the FAA, could be effectively installed in a military environment without extensive testing. This assumption proved to be fatal as the T-3 experienced numerous mechanical failures and significant difficulties integrating into the military environment.

B. LIMITED TESTING OF THE SLINGSBY T-3A FIRELY

As a COTS acquisition, the T-3 was only subjected to limited operational testing. As with other streamlined acquisitions, OT&E of the T-3 was abbreviated and many requirements were explicitly waived. Included in this waiver was the elimination of "missionized OT&E" prior to the purchase decision. Streamlined acquisition further increased program risk by not requiring an early operational assessment. As a result, the T-3 acquisition proceeded to full production and fielding without supporting OT&E. According to the Air Force, "this was considered an acceptable risk . . . due to the COTS/NDI nature of the T-3A." While most T-3 operational test requirements were waived, the Air Force did attempt limited testing.

In 1993, the T-3 underwent eight days of qualification operational testing. Not only was this initial testing extraordinarily limited in duration, but the Air Force allowed Slingsby to conduct the tests independently without AFOTEC involvement. This decision ignored the usual OT&E procedure of employing typical military operators. Furthermore, the principle of conducting tests in a realistic operational environment was ignored as aerial tests were only performed at Hondo Air Force Base in Texas and not at the Air Force Academy in Colorado. The limited testing of the T-3 was "operational" in name only as it completely failed to follow many of the fundamental principles of OT&E.

---

110 Baker, supra note 85, at 55.
112 Id.
113 Id. at 14.
114 Id.
115 Baker, supra note 85, at 55.
116 Id.
117 Id.
Though the Air Force attempted to involve AFOTEC in later testing, these tests were ultimately cancelled due to late delivery.\(^1\) Eventually, funding for later OT&E was denied entirely.\(^2\) The extraordinarily limited testing of the T-3 was seriously hindered by the inherent conflicts and data limitations of contractor-conducted testing.\(^3\) Additionally, the Air Force completely failed to test the aircraft in one of its intended environments.\(^4\) More thorough testing of the T-3 might have exposed the maelstrom of problems it would soon encounter.

Despite the readily apparent inadequacy of operational testing, the Air Force accepted delivery of the first of 110 T-3s in February 1994.\(^5\) The acquisition costs of this fleet totaled only $32 million—well below the financial thresholds mandating operational testing.\(^6\) Fifty-three planes were immediately sent to the Air Force Academy in Colorado with the remainder going to Hondo in Texas.\(^7\) At the Academy, the limited operational testing of the T-3 would soon collide with real-world military operations.

C. THE ULTIMATE CONSEQUENCES OF T-3A FIREFLY OPERATIONS

Problems with the T-3A Firefly became apparent soon after Academy aerial operations began. By 1996, Academy maintenance teams were making round-the-clock repairs, but were unable to assuage the constant problems.\(^8\) At the Academy, T-3 engines failed sixty-six times during takeoff or landing, forcing officials to ground fifty-seven planes on ten occasions.\(^9\) These problems were attributed to deficiencies with the T-3’s engine, fuel system, and brakes.\(^10\) Despite the efforts of Academy maintenance personnel, the problems with the T-3 proved to be insurmountable.\(^11\)

\(^{11}\) Baker, supra note 85, at 55.

\(^{12}\) SEC’Y OF THE AIR FORCE INSPECTOR GEN., supra note 83, at 17.

\(^{13}\) Baker, supra note 85, at 55; see also T-3A Firefly, supra note 107.

\(^{14}\) Baker, supra note 85, at 55.

\(^{15}\) SEC’Y OF THE AIR FORCE INSPECTOR GEN., supra note 83, at 4, 23.

\(^{16}\) T-3A Firefly, supra note 107.

\(^{17}\) Id.

\(^{18}\) Thompson, supra note 81.

\(^{19}\) Id.

\(^{20}\) T-3A Firefly, supra note 107.

\(^{21}\) Id.

\(^{22}\) Thompson, supra note 81.
The first fatal crash occurred within a year of Academy cadets flying the T-3. In twenty-seven months, three Academy planes crashed—killing three cadets and their instructors. On February 22, 1995, Cadet Mark Dostal and Captain Dan Fischer died in the first T-3 crash. A subsequent Air Force investigation concluded Cadet Dostal inadvertently put the plane in a spin, and Captain Fischer was unable to recover due to inadequate training. On September 30, 1996, Cadet Dennis Rando and Captain Clay Smith perished in the second fatal accident. Nine months later, Cadet Pace Weber and Captain Glen Comeaux died when their engine failed at 500 feet. Immediately following this third fatal crash, T-3 operations were suspended.

During this suspension, the Air Force investigated its acquisition of the T-3. The investigation revealed that Slingsby recommended 119 different changes to its COTS aircraft during its initial testing. Though it remains uncertain how many of these changes were incorporated into the T-3, at a minimum, the vast quantity of recommended changes illustrates the inherent difficulties of adapting commercial aircraft to a military environment.

Furthermore, the systemic impact of these modifications was not thoroughly tested. Former T-3 command instructor pilot Captain Pat Derock noted that no one seemed to “know what testing went into all those different changes.” The T-3 stands as a paradigmatic example of the integration, functional, and data problems inherent to COTS acquisitions. Unfortunately for the Air Force and its young pilots, these issues produced fatal results.

The investigation conducted during the suspension ultimately recommended returning the T-3 to an earlier acquisition

---

130 Id.
131 Id.
132 Id.
133 Id.
134 Id.
135 Id.
136 Id.
137 Id.
138 Id.
139 Id.
140 Id.
141 Id.
The committee suggested that full OT&E be conducted—including operations at the Air Force Academy and the evaluation of instructor pilot training. These recommendations were never enacted because the Air Force ultimately decided to terminate aerobatic flight screening and return the program to a focus on basic flight principles.

Deeming the T-3 acquisition a complete failure, the Air Force dropped the plane from its inventory in October 1999 without ever resuming operations. The Air Force blamed the T-3’s failure on its inappropriate COTS acquisition strategy. The Air Force concluded that the “substantial modifications” required to operate the T-3 in a military environment negated its COTS character and warranted full OT&E. Following unsuccessful attempts to individually sell its 110 Fireflies or sell them for parts, the Air Force ultimately disposed of the entire fleet as scrap metal. The disastrous acquisition of the T-3A Firefly resulted in a nearly $40 million loss, irreparably damaged the Air Force’s reputation, and cost six young pilots their lives.

D. THE ROLE OF OPERATIONAL TESTING IN AVOIDING THE TRAGEDY

The 1998 investigation uncovered a variety of different reasons for the aircraft’s failure. These oversights include operational, geographic, and personnel issues specific to the T-3. Ultimately, assigning complete or even substantial blame to any individual factor is irrelevant. Full-scale operational testing would have addressed each of these issues and almost certainly identified the problems inherent to T-3 operations.

The 1998 investigation pointed to the extreme air traffic at the Academy as one factor leading to the T-3’s failure. The report noted that “the Academy Airfield is an extraordinarily busy facility.” In 1996 alone, the Airfield supported 173,078

---

142 Sec’y of the Air Force Inspector Gen., supra note 83, at 5, 27.
143 Id.
144 T-3A Firefly, supra note 107.
145 Id.
146 Baker, supra note 85, at 56.
147 Id.
149 Baker, supra note 85, at 53.
150 See, e.g., Sec’y of the Air Force Inspector Gen., supra note 83.
151 Id. at 32.
152 Id.
aerial missions, while the pilot training bases of Sheppard and Laughlin respectively supported only 70,883 and 49,878 similar activities.\textsuperscript{153} The report also noted that “[b]y contrast, the level of non-military operations in the vicinity of Hondo is extremely low.”\textsuperscript{154} The high level of aerial traffic at the Academy was exacerbated by the proximity of the Colorado Springs municipal airport.\textsuperscript{155}

The intensity of aerial activities at the Academy complicated local T-3 operations.\textsuperscript{156} While this problem certainly impacted the Academy crashes, it could have been exposed during full OT&E. OT&E entails assessment under realistic operational conditions and would have certainly included aerial tests at the Academy Airfield.\textsuperscript{157} The Air Force ignored this fundamental principle by conducting tests solely at Hondo.\textsuperscript{158} Operational tests at the Academy could have revealed the impact of heightened aerial traffic on T-3 operations.

In addition to the heightened air traffic, the Academy also presented unique geographic considerations.\textsuperscript{159} The Academy’s elevation of 6,572 feet both limited operations and negatively impacted the T-3’s performance.\textsuperscript{160} The report noted “the Academy’s vertical airspace is significantly constrained by geography and topography.”\textsuperscript{161} The vertical airspace limitation effectively forced pilots to perform aerial operations at a lower above-ground altitude or face the T-3’s altitude-created mechanical problems.\textsuperscript{162} The report noted that the English-manufactured engine was not powerful enough to do aerobatic maneuvers in “the thin Rocky Mountain air.”\textsuperscript{163} The report also noted that no accidents occurred at Hondo and that many pilots believed the T-3 simply flew better “in the lower, and heavier, Texas air.”\textsuperscript{164} Once again operational tests conducted in a realistic environment could have exposed this problem earlier. Test flights conducted in the thin air of the Rocky Mountains would have

\textsuperscript{153} Id. at 31.
\textsuperscript{154} Id. at 33.
\textsuperscript{155} Id. at 32.
\textsuperscript{156} Id.
\textsuperscript{157} See id. at 13.
\textsuperscript{158} Id. at 16.
\textsuperscript{159} Id. at 30.
\textsuperscript{160} Id. at 33.
\textsuperscript{161} Id.
\textsuperscript{162} See id. at 33–34.
\textsuperscript{163} Thompson, supra note 81.
\textsuperscript{164} Id.
certainly uncovered the aircraft's reduced performance. Consequently, operational testing would have uncovered the impact of the Academy's elevation on aerial operations long before fielding, which could have avoided the T-3 tragedy.

Finally, the report concluded several personnel limitations contributed to the T-3's failure. First, the report noted the "marked contrast" between the student environments at the Academy and Hondo. The intense distractions of cadet-life at the Academy often prevented quality crew rest for student pilots. Conversely, Hondo students were temporarily assigned to pilot duties and were totally devoted to flying. Additionally, Academy instructor pilots were frequently academic professors assigned other military duties. Regarding personnel matters, General McPeak noted that "[m]aybe if you'd had three fighter pilots in there instead of three C-141 pilots you wouldn't have had the same result." While General McPeak's conclusion appears extraordinarily insensitive, it also fails to adequately address the T-3's shortcomings. Assigning blame to the student environment, instructor distraction, or inadequate instructor selection is irrelevant because early OT&E could have addressed each of these problems.

Operational testing conducted by typical military users in a realistic environment could have exposed each of the T-3 shortcomings identified in the Air Force investigation. Operational testing is designed to identify these types of problems and frequently exposes a myriad of training, personnel, doctrinal, and systemic deficiencies. Such early operational testing of the T-3 could have easily identified its air traffic, geographic, and personnel issues prior to its tragic fielding.

---

165 Sec'y of the Air Force Inspector Gen., supra note 83, at 35-36.
166 Id. at 35.
167 Id.
168 Id. at 42.
169 Id. at 35.
170 Thompson, supra note 81.
171 Baker, supra note 85, at 55.
172 See Harlow & Piegzik, supra note 27 (demonstrating that AFOTECH operational tests identify a variety of suitability deficiencies beyond mere system performance).
V. CONTINUED FAILURES OF OTHER COTS AIRCRAFT ACQUISITIONS

While the T-3 constituted a complete failure to purchase an entire COTS aircraft, the Air Force has also struggled to integrate COTS components into existing aircraft. While not resulting in fatalities, the T-45, C-5, and C-130J demonstrate the difficulties of integrating COTS aircraft components and the resulting financial losses. Unfortunately, the T-3 was neither the first, nor the last, COTS acquisition failure. Concurrent with the T-3 acquisition, the Air Force attempted a non-developmen-
tal purchase of the T-45. The T-45 Joint Primary Aircraft Training System (JPATS) is a trainer providing intermediate strike training. Similar to the T-3, the Air Force purchased an existing foreign aircraft to become the T-45. Though JPATS was designated a streamlined acquisition, significant handling problems delayed fielding by ten years and required a 90% in-
crease in testing. Once again, FAA certification was relied upon, but the RAND study concluded this certification only demonstrated basic airworthiness and not compliance with stringent military requirements. While employing an acquisition strategy identical to the T-3, the JPATS encountered similar difficulties resulting in a decade-long fielding delay and nearly double the projected expenditures.

More recently, the Air Force instituted component upgrades of both the C-5 and C-130J. In the C-5 Avionics Modernization Program, the Air Force encountered significant problems integrating FAA certified components into an existing aircraft. This program attempted to replace the C-5's analog cockpit instruments with digital displays and equipment. The detachment commander, Major Chris Dobb, commented, "[w]e had a lot of problems moving avionics from the lab to the aircraft." The on-site chief engineer, Wade Smith, bluntly de-

---

173 Fox et al., supra note 15, at 36.
174 Id. at 37.
175 Id. at 36.
176 Id. at 37.
177 Id.
178 Id. at 36-37.
180 Id.
181 Id.
182 Id.
scribed these problems and succinctly summarized the problems of integrating COTS components into an existing aircraft:

Commercial-off-the-shelf is a myth . . . . A commercial airliner takes off, goes to altitude, cruises, descends, and lands. A commercial system is not designed to fly to altitude, descend, and rendezvous with a tanker, descend, and ingress at 300 feet altitude. A flight management system for a commercial airliner could choke on that mission profile.185

Similar problems were encountered with the C-130J. The J Model implemented software upgrades to the C-130 series of medium range transport aircraft.186 C-130J testing was only used to supplement FAA certification, and iterative software modifications led to cascading problems causing unanticipated delays and expenditures.187

The C-5 modernization and C-130 upgrade demonstrate the substantial problems encountered while integrating COTS components into an existing aircraft. While these problems did not cause fatalities, they still resulted in substantial delays and increased expenditures. The continued pressure for faster and cheaper military development reveals that even after the T-3, the myth of a COTS military aircraft persists.

VI. RETURNING TO THE T-3: ULTIMATE LIABILITY

While the ultimate legal resolution of the T-3 tragedy illustrates the fundamental flaws of COTS aircraft acquisitions, it also reveals a possible solution. As a threshold matter, the Feres Doctrine prevents military personnel from suing the government for injuries sustained in the course of their service.188 Feres immunity aims to protect the special relationship between soldiers and avoid the extreme results tort liability would have on military discipline.189 While the estates of the deceased T-3 pilots were prohibited from suing the United States, the nature of the T-3 acquisition did not shield Slingsby Aviation from liability.

In Boyle v. United Technologies Corp., the Supreme Court established the parameters of the government-contractor defense.190

185 Id.
186 FOX ET AL., supra note 15, at 38.
187 Id.
188 Id.
191 Id.
As an initial matter, the Court rejected the extension of the *Feres* Doctrine to military contractors. The Court feared such an extension would establish a complete shield against contractor liability even for stock purchases. From this position, the Court established the requisites for denying contractor liability in military accidents.

From a general standpoint, the government-contractor defense described in *Boyle* shields contractors from liability resulting from discretionary design decisions made by the government. According to *Boyle*, contractor liability cannot be imposed for defects in military equipment when: (1) the government approved reasonably precise specifications; (2) the product conformed to those specifications; and (3) the contractor warned the government about potential dangers known only to the contractor. According to the Court, the first two requirements ensure government consideration of the specific feature, while the third condition encourages contractors to disclose information relevant to product liability.

In *Boyle*, the Court also offered a specific example when contractor liability is appropriate. The Court noted that if the government purchased stock helicopters without substantial modifications, the contractor would bear ultimate tort liability. The *Boyle* test for contractor liability and the specific example provided by the Court left little doubt that Slingsby would be liable for the T-3 tragedy.

In 2001, a federal jury imposed liability on Slingsby for the third fatal accident and awarded Cadet Pace Weber's parents $4 million for the loss of their son. In *Weber v. Slingsby Aviation Ltd.*, the court applied the *Boyle* test and concluded the government did not exercise the necessary discretion to invoke the government-contractor defense. The court found that Slingsby was unable to provide any evidence that the T-3 was designed in accordance with Air Force-provided specifications or that the
Air Force exercised any discretion in directing its design.\textsuperscript{197} Shortly after the Weber decision, Cadet Dennis Rando’s parents reached a confidential settlement with Slingsby.\textsuperscript{198}

The interaction of Boyle and Feres reveal that two basic mechanisms protect American soldiers from the negative consequences of ill-suited and ineffective government purchases. The Feres Doctrine implicitly rests on the notion that some mechanism protects American soldiers from harm caused by custom-built military equipment. This mechanism is operational test and evaluation. Conversely, when injuries result from a COTS acquisition, contractor liability compensates military personnel for their injuries. The T-3 tragedy illustrates that a gap exists between these two doctrines. This gap occurs when a COTS acquisition failure has fatal consequences, as it did with the T-3.

Though the failure of certain COTS acquisitions can have fatal results, operational testing does not provide military personnel with an added layer of protection. In the event of fatal COTS acquisition failures, contractor liability only serves to compensate the survivors and provide long-term incentives for product improvements. Nothing protects the lives of military personnel using these products. This result is unacceptable.

VII. SOLUTION: EXPAND MANDATORY OT&E TO “PERILOUS” NONCOMBAT COTS ACQUISITIONS SUCH AS TRAINER AIRCRAFT

A relatively simple statutory modification to 10 U.S.C. § 2399 would expand mandatory OT&E to products like the T-3 and could prevent similar tragedies. Though mandating OT&E for previously exempt COTS products would certainly increase programmatic expenditures, several avenues exist to mitigate these costs. Cost-effective, mandatory OT&E provides a viable and beneficial mechanism to prevent disastrous acquisitions like the T-3 in the future.

A. CHANGE MECHANISM: STATUTORY CHANGE

American soldiers using potentially lethal products deserve protection beyond the mere financial compensation provided by contractor liability. The T-3, JPATS, C-5, and C-130J demonstrate that even minor modifications can alter the military-specific performance of COTS aircraft. Operational testing

\textsuperscript{197} Id.

\textsuperscript{198} British Training Aircraft Maker Settles, supra note 195.
guarantees evaluation of the full impact of these modifications. Operational testing also ensures evaluation of these products by future users in an operationally realistic environment. The motivations of test personnel spring from future personal use, as well as use by friends and colleagues. The personal ownership of the risks inherent to these products provides a motivation above and beyond the financial incentive of government contractors. This motivation can ensure that perilous COTS products, such as the T-3, are both effective and suitable prior to fielding.

Currently, operational test and evaluation is only required for major defense acquisition programs.\footnote{10 U.S.C. § 2399(a)(1) (2006).} This statute should be modified with the insertion of the phrase "or potentially perilous" after "major" in 10 U.S.C. § 2399(a)(1). This phrase would encapsulate noncombat programs that are not expensive enough to rise to the level of major according to the statute. Trainer aircraft are the perfect example of an acquisition that is neither major nor combat-related and therefore exempt from mandatory operational testing. Essentially, these programs slip through the OT&E cracks of Title X. Though not subjected to mandatory operational testing these programs have proven to be lethal to their soldier-operators and have cost the government millions of dollars.

While the Air Force and the DOD learned valuable lessons from the failure of the T-3, subsequent COTS acquisitions demonstrate the fickle nature of these lessons. Political pressures will certainly rise again and suggest saving tax dollars through similar COTS acquisitions. However, avoiding the expenses associated with operational testing does not justify exposing soldier-operators to fatally defective products. Military pilots should never again be exposed to an aircraft that is ineffective in its intended environment and unsuitable to be flown by the very pilots for whom it was designed. A statutory change mandating OT&E of these products would cement the lessons of the T-3 for future DOD and political leaders.

B. Cost-Effective Implementation of Operational Testing

Increased operational testing will certainly increase acquisition expenditures; however, this expense is not insurmountable. Operational testing is very expensive and critics will certainly note that mandatory OT&E would eliminate most of the finan-
cial benefits of COTS acquisitions. As noted earlier, operational testing accounts for over one-fifth of the developmental costs for fixed-wing aircraft.\textsuperscript{200} Though mandatory OT&E would expand these expenses to previously exempt acquisitions, several mechanisms exist to reduce these costs. An expansion of the government-contractor defense, implementation of innovative test techniques, and the general savings reaped from fewer program failures would allow the government to recoup a large portion of these expenses.

First, the government-contractor defense should be expanded to preclude liability for these "perilous" products. Mandatory OT&E combined with an expansion of the contractor defense shifts the responsibility for program failure from the contractor to the government. While the \textit{Feres} doctrine prevents military personnel from suing the government,\textsuperscript{201} this prohibition is supplemented by OT&E. Operational testing provides a superior mechanism to expose design deficiencies prior to fielding. The advent of mandatory OT&E of these products would effectively reduce the possibility of operational accidents.

Additionally, an expansion of the government-contractor defense would reduce contract-related expenses. Military suppliers typically pass the cost of accidents to the government through cost overrun provisions.\textsuperscript{202} Furthermore, the price of liability insurance is passed to the government either in the specific contract or through higher prices in later contracts.\textsuperscript{203} Consequently, the expansion of the contractor defense would reduce up-front contract expenses.

The test and evaluation community consists of practioner-academics constantly updating and improving test techniques. Within this community, a variety of testing innovations have been developed and implemented.\textsuperscript{204} OT&E experts have also suggested implementing iterative, mathematical techniques that can improve the effectiveness of OT&E and ultimately streamline testing.\textsuperscript{205} Mathematical and analytical techniques em-

\textsuperscript{200} \textit{Fox et al.}, \textit{supra} note 15, at xv n.1.

\textsuperscript{201} \textit{Feres v. United States}, 340 U.S. 135, 146 (1950).

\textsuperscript{202} \textit{McKay v. Rockwell Int'l Corp.}, 704 F.2d 444, 449 (9th Cir. 1983).

\textsuperscript{203} \textit{Id.}

\textsuperscript{204} \textit{See} Christopher L. Harlow & Santa Falcone, \textit{A Correlated Strategic Guide to Software Testing}, \textit{CROSSTALK}, July 2005, at 18–21 (discussing various emerging practices and suggesting a new cost-effective iterative, mathematical model).

\textsuperscript{205} \textit{Id.}
ployed during testing can be used to better focus OT&E. This focus reduces the necessary time and expenditures without diminishing test quality. These relatively inexpensive progressive models can be used to improve test efficiency and reduce the expenditures associated with OT&E.

Finally, mandatory OT&E creates obvious savings opportunities and could easily pay for itself. Early deficiency exposure in mandatory OT&E could prevent calamitous purchases such as the T-3. In the case of the T-3, the Air Force purchased its fleet for nearly $40 million only to dispose of the planes as scrap metal and sacrifice virtually its entire investment. Avoiding similar financial disasters would allow this statutory modification to pay for itself.

The relative ease in which test costs can be reduced further justifies mandatory OT&E for COTS aircraft acquisitions. American soldiers deserve the maximum protection possible from the potentially fatal consequences caused by the failure of these programs. Mandatory operational test and evaluation would provide this protection. A statutory change mandating OT&E for potentially perilous products would memorialize the lessons of the T-3A Firefly and take a significant step towards preventing similar tragedies in the future.

206 *Id.*
207 *Id.* at 21.
208 See *id*.
209 Officials Announce T-3A Firefly Final Disposition, supra note 148; see also Baker, *supra* note 85, at 53.