INSTITUTE FOR DEFENSE ANALYSES

Independent Assessment of Missile Defense Agency Processes for Flight Test Planning and Execution

Joseph T. Buontempo, Project Leader

John S. Hong, Deputy Project Leader

Matthew R. Girardi

Jeremy A. Teichman

Robert V. Uy

February 2022

IDA Publication P-30853

Log: H 2021-000403

INSTITUTE FOR DEFENSE ANALYSES
730 East Glebe Road
Alexandria, Virginia 22305

IDA

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication

This work was conducted by the IDA Systems and Analyses Center under contract HQ0034-19-D-0001, Project DB-1-4959, "Missile Defense Agency (MDA) Processes for Ballistic Missile Defense System (BMDS) Flight Test Planning and Execution" for the DB / MDA / Missile Defense Agency. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization. Approved for public release; distribution is unlimited.

Acknowledgments

The authors would like to thank the IDA committee, Dr. Stephen Ouellette (chair), Dr. Dawn C. Loper, Dr. Michael L. Luhman, Dr. Angelo J. Signoracci and Dr. Kevin Tian for providing technical review of this effort.

For More Information

Joseph T. Buontempo, Project Leader
jbuontem@ida.org , (703) 845-2349

Stephen M. Ouellette Director, SED
souellet@ida.org, (703) 845-2443

Copyright Notice

Rigorous Analysis | Trusted Expertise | Service to the Nation

INSTITUTE FOR DEFENSE ANALYSES

IDA Document P-30853

Independent Assessment of Missile Defense Agency Processes for Flight Test Planning and Execution

Joseph T. Buontempo, Project Leader

John S. Hong, Deputy Project Leader

Matthew R. Girardi

Jeremy A. Teichman

Robert V. Uy

Executive Summary

The Government Accountability Office (GAO) completed its report, *Missile Defense: Assessment of Testing Approach Needed as Delays and Changes Persist*, in July 2020 (GAO Report 20-432). One of the report's main findings is that

> "[Missile Defense Agency] MDA did not fully execute its fiscal year 2019 flight testing, continuing a decade-long trend in which MDA has been unable to achieve its fiscal year flight testing as scheduled. Although MDA revised its approach to developing its annual test plan in 2009 to ensure the test plan was executable, over the past decade MDA has only been able to conduct 37 percent of its baseline fiscal year testing as originally planned due to various reasons including developmental delays, range and target availability, or changing test objectives. In addition, MDA has not conducted an assessment to determine whether its current process for developing and executing its annual test plan could be improved to help ensure its executability. Without an independent assessment, MDA will continue down the same path, increasing the risk of the same outcomes from the past decade—less testing than originally planned, resulting in less data to demonstrate and validate capabilities."

The GAO report then recommended that "MDA ensure an independent assessment is conducted of its process for developing and executing its annual [U.S. Ballistic Missile Defense System] BMDS flight test plan." The Department of Defense (DoD) concurred with this recommendation, and MDA asked the Institute for Defense Analyses (IDA) to conduct the assessment.

The scope of IDA's study included the BMDS flight test (FT) program and its role in characterizing the system's specific capabilities and limitations, demonstrating the overall BMDS capability to allies and adversaries, deterring adversaries' use of ballistic missiles, building warfighter confidence, and validating and verifying the BMDS models and simulations. Other aspects of the BMDS test program – such as ground tests, the use of models and simulations, exercises and wargames – were outside the scope of this assessment.[1]

IDA's overarching approach to conducting the study began by reviewing relevant documentation and conducting interviews regarding MDA's annual BMDS FT program.

[1] It is worth noting that examining the relations and timing between ground tests and flight tests could potentially help improve the overall BMDS development process, especially with regard to timely verification, validation, and accreditation of ground test models using flight test data.

For the interviews, IDA engaged in discussions with government officials from MDA and other relevant DoD stakeholder organizations to hear their perspectives and learn from their experiences. These interviews were not for attribution. The information was analyzed and common trends and issues were identified. Finally, the study team formulated findings and recommendations.

IDA found the FT process MDA actually executes largely follows the documentation, with stakeholders generally speaking positively of the process. In addition, the team found that the heavily iterative nature of generating an IMTP (Integrated Master Test Plan) currently requires repeated instances of substantial input and coordination from numerous internal and external stakeholders.[2] Delayed coordination can lead to disruption of the test planning process, which was one of the more common frustrations voiced by interviewees. IDA also observed that, although operational testers external to MDA are engaged early in the overall test planning process (i.e., IMTP), they are often involved late in the detailed planning process for a specific test.

IDA was able to replicate GAO's estimate (37 percent) for MDA's FT execution rate when counting all FTs over all IMTPs. This estimate captures both execution and schedule churn. When the team considered FT changes within two years of the original best estimated test date, it appears MDA is executing 54 percent of its planned tests. In addition, IDA notes that the number of FTs per year and the total test-related spending appear to be growing by approximately the same percentage. But by some measures, the systems under test and the scenarios designed to test them are getting more complex.

GAO's report listed several possible root causes for FT delays and cancellations. IDA found there were also a number of additional causes. Any related evaluation of MDA should account for the causes of delay, the degree to which they could or should have been avoided, and their consequences. Using an IDA-developed taxonomy, the team found that about half of FT delays, modifications, and deletions are characterized as intentional and the other half are unintentional. About a third of all delays, modifications, and deletions are due to factors external to MDA. Although IDA found no dominant reason for test schedule changes, programmatic reasons are cited for the largest fraction of unintentional internal FT changes, and that this category should be further developed to more clearly identify the causes. Cascading delays, where one delayed test postpones successive tests in a chain reaction, might be avoidable with lower test cadences. However, trading lower test cadence for lower probability of cascading delays does not provide an obvious net benefit to MDA, and delays might sometimes be the lesser price to pay. Where cascading schedule disruptions are a possibility, MDA should certainly account for possible failures or delays and attempt to build resilience into the schedule.

---

[2] External stakeholders include DOT&E, DTE&A, AFOTEC, ATEC, COTF, JFCC-IMD, NORTHCOM, and BMDS OTA.

The BMDS FT program serves a number of goals: (1) Demonstrating BMDS Capability, (2) Characterizing BMDS Capabilities and Limitations, (3) Deterrence and Warfighter Confidence, and (4) Support of modeling and simulations (M&S) Verification, Validation, and Accreditation (VV&A). These goals often conflict with one another, and all must be served by the same highly constrained FT program. IDA examined the interplay of causes and consequences of FT cancellations, delays, and changes as well as other effects found in the review of MDA's FT processes.

When demonstrating BMDS capability and verifying functionality to support Phase Implementation Plan (PIP) increments, cancellations and delays in flight-test-based system verification often force MDA to devolve to lower-confidence verification methods. IDA found that such downgrading is not accompanied by traceable justifications and as a result, consequences of FT cancellations or delays on system demonstration are not prominently observable. MDA also has challenges associated with its current process for tracking assessment and verification requirements, and progress toward verification goals. This process has need of a feedback loop to circle back after testing to confirm which verification goals were satisfied and the means by which they were satisfied to track progress and project future progress. Historically, M&S was a driver of MDA's FT program. Due to the inability to exhaustively test the function of the BMDS in FTs, MDA established a strategy to efficiently use the limited FT program for VV&A of the M&S. IDA found that less emphasis is now placed on the use of FT data for M&S VV&A. Delays and cancellations of FTs as well as de-scoped tests further diminish or delay the validation of M&S, encouraging the use of M&S without full VV&A and exacerbating the dependence of system acceptance decisions on unaccredited modeling.

Based on IDA's assessment of the frequency of FT delays and cancellations, causes of delays, the degree to which they could or should have been avoided, and their consequences, IDA recommends that MDA (1) conduct a trade study to assess the costs and benefits of adjusting pre-mission test margin (with unchanged FT cadence), (2) review the IMTP generation process and reconsider if stakeholder coordination is truly required in each of the IMTP process steps and streamline the process as appropriate, (3) use operational testers early, during the first year of detailed FT planning, (4) develop and apply a taxonomy to monitor causes of FT cancellation/delays and use the results to improve FT planning and execution, (5) develop a single-source traceable end-to-end mapping of assessment objectives for tests to evaluate progress toward meeting verification goals (the process should include reviving the use of Critical Engagement Conditions/Empirical Measurement Events (CEC/EMEs)[3] or similar but more comprehensive measures to manage the accomplishment of FT goals and include a

[3] CEC/EMEs were high-profile metrics and as such were encouraged to be kept to a manageable number. Similar but lower-profile measures could potentially be much more comprehensive.

feedback loop that circles back after testing to confirm which verification goals were satisfied and the means by which they were satisfied), and (6) establish a formal tracking procedure to better understand confidence down-grading when relying on lower-confidence verification methods (i.e., non-FT based).

Contents

Tables

Figures

1. Introduction

The mission of the Missile Defense Agency (MDA) is "to develop and deploy a layered Missile Defense System to defend the United States, its deployed forces, allies, and friends from missile attacks in all phases of flight."[4] Generally, MDA's focus has been defense against ballistic missiles of all types, though more recently it has begun to examine the possibility of defenses against other missile threats such as cruise missiles and hypersonic weapons. Currently, its major responsibility is developing and deploying the U.S. Ballistic Missile Defense System (BMDS), a geographically distributed system-of-systems that includes sensors, interceptors, and command and control systems. For example, the Ground-based Midcourse Defense (GMD) system and supporting surface-based radars – Cobra Dane, Upgraded Early Warning Radars (UEWRs), Sea-based X-band radar (SBX), and Long Range Discrimination Radar (LRDR) – are used for Homeland Defense against intercontinental ballistic missiles (ICBMs) and intermediate range ballistic missiles (IRBMs). Theater Defense systems – employed against IRBMs, medium range ballistic missiles (MRBMs), and short range ballistic missiles (SRBMs) – include Aegis BMD ships, Aegis Ashore, Patriot, Terminal High Altitude Area Defense (THAAD), and TPY-2 radars.[5] The BMDS also includes space-based sensors (e.g., Space-Based Infrared System (SBIRS)) and the Command and Control, Battle Management, Communications (C2BMC) system.

A. Overview of the Missile Defense Agency's Ballistic Missile Defense System Flight Test Program[6]

MDA's program for testing the BMDS is intended to collect "critical data to demonstrate the operational effectiveness, suitability, and survivability" of the BMDS and to contribute "to U.S. non-proliferation goals by sending a very credible message to the international community on our ability to defeat ballistic missiles in flight, thus reducing their value to potential adversaries." The MDA/Directorate for Test (MDA/DT), located primarily in Huntsville, Alabama, and Colorado Springs, Colorado, "executes BMDS test

[4] https://www.mda.mil/about/mission.html

[5] The TPY-2 radar can also be used for Homeland Defense.

[6] Section A. captures MDA's high-level representation and description of its BMDS Test Program, and, unless otherwise noted, the quotes in this section come from the MDA Fact Sheet: Ballistic Missile Defense Testing, September 2020 (20-MDA-10590).

policy, manages the BMDS Test Baseline, and provides programmatic and technical direction and oversight of the test program and test resources.”

According to MDA, BMDS testing “is a continuous, evolutionary process that encompasses both developmental and operational activities.” The process starts with “testing of individual system elements and components and progresses to end-to-end testing of the integrated system,” along the way adding “increasingly challenging objectives” and becoming “more operationally realistic.” Because BMDS testing is time consuming and expensive, MDA also relies on models and simulations “to assess system configurations, engagement conditions, and target phenomena” and to “allow repeated assessments of performance and provide a statistical determination of effectiveness.”

To collect “essential data to validate the accuracy of models and simulations,” MDA employs both FTs and ground tests:

- “Flight tests use assets in their operational configuration against an inventory of targets to assess all aspects of BMDS performance in a variety of flight test regimes.”
- “Ground tests combine element hardware-in-the-loop, digital representations, high fidelity threat simulations, and operational assets to test BMDS capabilities across a wide range of threats and environments that cannot be affordably replicated in flight tests.”

In addition to FTs, ground tests, and models and simulations, the BMDS test program employs exercises and wargames to “support Joint Staff, Combatant Commanders, Service Components, and Allies in preparation of concepts of operations; tactics, techniques, and procedures; doctrine; and requisite training on current and evolving BMDS capabilities.”

MDA plans and conducts BMDS tests “in partnership with the Director of Operational Test and Evaluation, Director of Developmental Test and Evaluation, and with the Army, Navy, and Air Force Operational Test Agencies to embed operational test and warfighter requirements in the test program.” The warfighters who operate the individual systems “ensure that tests use operational doctrine and real-world constraints while evaluating new concepts of operations and exercising tactics, techniques, and procedures.” Furthermore, MDA states that “testing provides warfighters with confidence in the basic design of the BMDS, its hit-to-kill effectiveness, and its inherent operational capability.”

The specific focus of this study was solely BMDS flight testing. MDA’s BMDS Flight Test Concept of Operations states that “MDA conducts flight tests to gather data, demonstrate BMDS capability, characterize system limitations, and verify/validate modeling and simulations (M&S). Flight test data support Element acquisition milestones, instills warfighter confidence, assesses operational capability, supports real-world missions, and/or updates M&S."[7]

B. GAO Assessment of BMDS Flight Testing

In July 2020, the Government Accountability Office (GAO) completed its 17th annual review of MDA's progress on the BMDS.[8] The purpose of this report was to address "(1) the extent to which MDA achieved fiscal year 2019 delivery and testing goals for BMDS elements, as stated in reported baselines; (2) the extent to which MDA's annual test plan is executable; and (3) broad challenges that could impact MDA's portfolio and any actions the agency has taken to address them." To conduct its assessments, GAO reviewed the planned fiscal year 2019 baselines, test plans since 2010, and other documentation and "assessed them against program and baseline reviews;" and interviewed officials from MDA and other Department of Defense (DoD) organizations. One of the report's main findings is:

> MDA did not fully execute its fiscal year 2019 flight testing, continuing a decade-long trend in which MDA has been unable to achieve its fiscal year flight testing as scheduled. Although MDA revised its approach to developing its annual test plan in 2009 to ensure the test plan was executable, over the past decade MDA has only been able to conduct 37 percent of its baseline fiscal year testing as originally planned due to various reasons including developmental delays, range and target availability, or changing test objectives. In addition, MDA has not conducted an assessment to determine whether its current process for developing and executing its annual test plan could be improved to help ensure its executability. Without an independent assessment, MDA will continue down the same path, increasing the risk of the same outcomes from the past decade—less testing than originally planned, resulting in less data to demonstrate and validate capabilities.

The GAO report recommended that "MDA ensure an independent assessment is conducted of its process for developing and executing its annual BMDS flight test plan." DoD concurred with this recommendation, and MDA asked the Institute for Defense Analyses (IDA) to conduct the assessment.

The GAO report also mentioned MDA's "success-oriented" approach. There are two meanings of success-oriented relevant to MDA, both of which will be addressed in this report. The first refers to the design of tests. For a test design to be success-oriented suggests that the test is constructed to have a high likelihood of a mission success (e.g.,

---

[7] "Ballistic Missile Defense System Flight Test Concept Of Operations (MDA Policies and Procedures for Execution)," MDA Directive 3000.10, June 20, 2019, 5.

[8] "Missile Defense: Assessment of Testing Approach Needed as Delays and Changes Persist," July 2020 (GAO Report 20-432).

successful hit-to-kill intercept). This type of approach avoids more complex or challenging intercept geometries, scenarios, etc., trading off stressing the system against demonstrating successful missions. The second meaning of success-oriented refers to the test program rather than the test design. A success-oriented program implies that scheduling and budgeting for test events presume success and nominal progress in all other events and activities. A success-oriented schedule may be less resilient to failures or delays, potentially leading to cascading effects of off-nominal events such as issue discoveries.

C. IDA Study

1. Objective and Research Areas

The objective of this study, sponsored by MDA, was to conduct an independent assessment of MDA's processes for developing and executing its BMDS FT program. The scope of the study included the BMDS FT program and its role in characterizing the system's specific capabilities and limitations, demonstrating the overall BMDS capability to allies and adversaries, deterring adversaries' use of ballistic missiles, building warfighter confidence, and validating and verifying the BMDS models and simulations. Other aspects of the BMDS test program – such as ground tests, the use of models and simulations, exercises and wargames – were outside the scope of this assessment.[9]

The study occurred over the second half of 2021 and included the following major milestones:

- May 17 Study initiation and kick-off meeting with MDA
- June 29 IDA progress update briefing to MDA
- August 27 IDA progress update briefing to MDA
- Nov 30 IDA briefing of results to MDA

The study team conducted research in three areas:

1. MDA's processes for planning, scheduling, and executing the BMDS FT program
   a. IDA reviewed how MDA develops test objectives (TOs) and plans, executes FTs, performs post-test analyses, and conducts change management.

[9] It is worth noting that examining the relations and timing between ground tests and flight tests could potentially help improve the overall BMDS development process, especially with regard to timely verification, validation, and accreditation of ground test models using flight test data.

   b. Based on this understanding of the FT processes, IDA assessed whether the guidance required for planning is adequate and complete and identified potential areas for improvement.

2. Causes for FT schedule revisions and instabilities

   a. IDA assessed the extent to which MDA has met its goals for past annual BMDS FT plans by examining the:

      1) Characterization of and reasons for test delays and cancellations

      2) Extent of schedule changes (including frequency and length)

      3) Type of knowledge gained from FTs

      4) Obstacles and challenges in meeting MDA's FT goals (e.g., developmental delays, system readiness, and range and target availability).

3. Implications of FT delays and cancellations

   a. IDA assessed the implication of FT delays and cancellation with respect to:

      1) Characterizing the system's specific capabilities and limitations

      2) Demonstrating the overall BMDS capability to allies and adversaries

      3) Building warfighter confidence

      4) Deterring an adversary's use of ballistic missiles

      5) Validating and verifying the BMDS models and simulations.

2. Tasks and Methodology

IDA's overarching approach to conducting the study began by reviewing relevant documentation and conducting interviews regarding MDA's annual BMDS FT program. The information was analyzed and common trends and issues were identified. Finally, the study team formulated findings and recommendations.

The documents and information examined included:

- National Defense Authorization Act for Fiscal Year 2010, Public Law 111-84, Sec 236, October 28, 2009.
- Plan for Integrated Master Test Plan and Budget Synchronization, MDA and Director, Operational Test and Evaluation (DOT&E), 24 March 2017.
- BMDS Change Management Process, MDA Manual 3500.01-M, 14 November, 2017.

- Missile Defense System Joint Engineering and Test Integrated Master Test Plan Development, MDA Instruction 3000.04-INS, 8 October, 2020.
- Missile Defense: Assessment of Testing Approach Needed as Delays and Changes Persist, GAO Report 20-432, July 2020.
- BMDS Flight Test Concept of Operations, MDA Directive 3000.10, 20 June, 2019.
- Director, Operational Test and Evaluation (DOT&E) FY19 and FY20 Annual Reports to Congress.
- Current and historical versions of MDA's:
  - Integrated Master Test Plan (IMTP)
  - IMTP Smart Books
  - Phase Implementation Plan (PIP)
  - Missile Defense Accountability Report (MDAR)
  - Test Objective Memoranda (TOM)
  - DoD Fiscal Year Budget Estimate Defense-Wide Justification Books

For the interviews, IDA engaged in discussions with government officials from MDA and other relevant DoD stakeholder organizations to hear their perspectives and learn from their experiences. These interviews were not for attribution. The study team began the interviews with high-level general questions and then followed up with more detailed discussions related to the interviewee's specific responsibilities. Representative high-level questions included:

- What are MDA's processes for developing, scheduling, and executing its FT program?
- What are the obstacles and challenges in meeting MDA's FT execution goals?
- How would you improve the process?

Over the months of July, August, and September, 2021, IDA conducted 28 interviews (via teleconference) with 52 government officials from the following stakeholder organizations:

- Missile Defense Agency Directorate for Test (DT), Directorate for Engineering (DE), and Director for Operations (DO)
- Office of the Director, Operational Test and Evaluation
- Developmental Test Evaluation & Assessments
- Service Operational test components

  - Air Force Operational Test and Evaluation Center (AFOTEC)
  - Commander, Army Test and Evaluation Command (ATEC)
  - Commander, Operational Test and Evaluation Force (COMOPTEVFOR)
  - BMDS Operational Test Agent (OTA)
- Joint Functional Component Command for Integrated Missile Defense (JFCC-IMD)
- U.S. Northern Command

3. Overview and Report Structure

The GAO report focused on the frequency of FT delays and cancellations. IDA was asked to review MDA's FT planning and execution processes to investigate possible connections to the delays and cancellations. There are five top-level important questions associated with IDA's task.

1. Does IDA's estimation of FT delays and cancellations match those of GAO?
2. If so, why are FTs so frequently delayed or cancelled (proximate causes)?
3. What are the consequences of the delays and cancellations?
4. What are the root causes of those FT delays and cancellations that should be avoided?
5. How can the root causes be remedied? In particular, how might the FT planning and execution process be modified to either reduce the frequency of delays and cancellations or mitigate the consequences?

Chapter 2 of this report covers IDA's review and assessment of MDA's processes for planning and executing its FT program and presents several potential improvements. Chapter 3 confirms GAO's quantitative estimate of the frequency of FT delays and discusses IDA's assessment and characterization of proximate causes of FT schedule revisions and instabilities. GAO's report lists a number of potential causes focusing on those more likely to be problematic. However, it was discovered that many of the delays were due to causes other than those listed by GAO. The degree to which the delays and cancellations represent a problem depends on their consequences, which is explored in Chapter 4. IDA's overarching findings and recommendations relating to root causes and remedies are provided in Chapter 5. Appendix A provides a list of the specific individuals the study team spoke to during its interviews. Appendix B describes issues associated with optimizing pre-flight schedule margin, one possible delay mitigation strategy. Finally, Appendix C defines acronyms and abbreviations used in this report.

(This page is intentionally blank.)

2. Assessment of MDA's Flight Test Processes

MDA's process for FT planning and execution has two components: a large-scope component that addresses FT planning across all of MDA's tests and a small-scope component that addresses the FT planning for individual test events.

FT planning begins at a high level, as TOs and test requirements (TRs) are introduced or updated. TOs provide the purpose of the test, while TRs provide the conditions required to satisfy the TO. These initial TOs and TRs then drive an iterative process in which TOs and TRs continue to mature and TDs, financial requirements, and schedule are developed. Although this development is always ongoing, the results are published twice per year in an IMTP. In this way, the IMTP is a plan, a document, and a baseline. The process for generating the IMTP is documented in MDA Instruction 3000.04.[10]

Planning for individual FTs begins roughly two years before the best estimated test date (BETD). Detailed test planning includes designing the scenario, identifying required hardware and software, and describing data collection and analysis plans. Mission executability is assessed and test teams are given test-specific training. Test execution is followed by determining if TOs were met and capturing lessons learned. The process for this test-event-specific planning is documented in MDA Directive 3000.10.[11]

The following sections summarize the large and small-scope components at a level that will provide sufficient context for understanding the rest of the report. The summaries are consistent with the content found in MDA's documentation[10,11] and a reader seeking additional details is encouraged to review these documents. Through interviews, IDA learned that the process described in MDA's documentation is largely consistent with how planning and execution actually take place. Thus, the reader can feel confident that both the formal documentation and the summaries below represent what occurs in practice.

A. Planning of Flight Tests Across MDA's Portfolio

MDA documents its plans for conducting all of its FTs in an IMTP, released twice each year. The IMTP includes TOs, TRs, TDs, and schedules. The process for generating the IMTP consists of four phases, as shown in Figure 2-1.

---

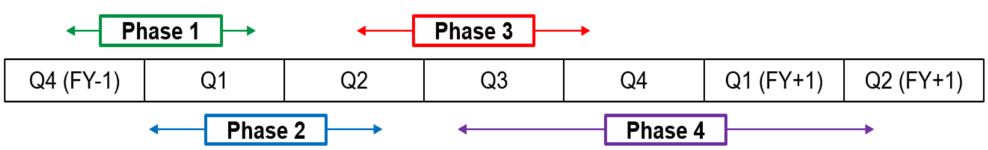
[10] Missile Defense System Joint Engineering and Test Integrated Master Test Plan Development; MDA Instruction 3000.04-INS; Missile Defense Agency; 2020 October 8.

[11] Ballistic Missile Defense System Flight Test Concept of Operations; MDA Directive 3000.10; Missile Defense Agency; 2019 June 20.

The primary MDA points of responsibility for IMTP development are the Directorate for Engineering (DE) and the Directorate for Test (DT). DE coordinates integration of TOs and TRs, while reviewing TDs. DT coordinates development of TDs, maintains an executable test baseline, and reviews TOs and TRs.

DE and DT are supported by the Directorate for Operations (DO) and Program Executive for Targets and Countermeasures (TC). DO ensures the budget and test program are synchronized and maintains the Missile Defense System cost model used to estimate test program costs. TC generates target solutions compatible with requirements and budget restrictions. DE, DT, DO, and TC work together to ensure synchronization of TOs, TRs, TDs, costs, and schedules.

Stakeholders in the IMTP generation process include internal and external partners, and are formally defined in MDA 3000.04: "MDA stakeholders include MDA directorates, programs, and functional areas that support Test Baseline development. External stakeholders include but are not limited to IMTP signatories, which include the following: Director, Operational Test and Evaluation, Deputy Director, Developmental Test, Evaluation and Assessment; Commanding General, Joint Functional Component Command for Integrated Missile Defense (representing the combatant commands); Commanding General, U.S. Army Test and Evaluation Command; Commander, Air Force Operational Test and Evaluation Center; Commander, Operational Test and Evaluation Force; and Commander, Joint Interoperability Test Command."



Phase	Overseen By	Significant Activities
1	Director for Engineering	Integrate initial test objectives and requirements.
2	Director for Test	Mature test requirements and objectives. Develop test designs and analyze costs.
3		Finalize test designs. Synchronize costs and schedules. Resolve outstanding issues. Publish IMTP version 0.
4		Update plan to account for resource changes. Publish IMTP version 1.

Figure 2-1. Timeline, Primary Point of Responsibility, and Significant Activities for each IMTP Generation Phase.[10]

1. Phase 1

Phase 1 begins with the end of the previous IMTP generation's Phase 3, nominally in August, and lasts roughly four months. Phase 1 is overseen by the DE.

The regular IMTP updates occur in response to changes in the elements and threats. The process begins with stakeholders providing DE with Element upgrades and responses to new or changing operational threats. These updates are accompanied by TO requests and the TRs that support those TOs. As necessary, stakeholders work alongside DE to form TOs and TRs, and develop notional TDs.

DE will review and refine newly submitted or updated TOs and TRs, ensuring consistency and compatibility with the existing technical baseline and schedule. Initial TOs, TRs, and notional TDs are then presented by DE at a series of reviews: first to the Systems Engineering Integration Council (SEIC) and then to the Executive Panel. Both the SEIC and the Executive Panel may request that the IMTP development team revise portions of the Phase 1 products. Once the SEIC and Executive Panel (IMTP signatories) give concurrence, Phase 1 concludes.

2. Phase 2

Phase 2 begins in the middle of Phase 1, nominally in October, and lasts for roughly five months. Phase 2 is overseen by the DT.

DE continues to revise TOs and TRs, obtaining consistency and compatibility with the technical baseline and schedule. Target requirements are assigned using threat-based information. These target requirements support initial cost estimates as well as enable TC to begin assessment of target solutions for test designs.

DT continues to mature TDs, coordinating with DE, TC, and stakeholders. TD planning includes potential trajectories, range safety, test resource placement, telemetry, truth coverage, target selection and feasibility, air launch target test planning, and ability to accomplish TOs and TRs. DT prioritizes and schedules TDs based on criteria such as first-of-its-kind capability testing, wider domestic and international community visibility, test resource impact, scheduling compatibility, cost, and risk.

Simultaneously to the DE and DT efforts, TC uses TOs, TRs, and design studies to explore executability, feasibility, and affordability issues pertaining to targets and countermeasures. TC proposes target and countermeasure solutions to satisfy DE-approved target requirements. DT provides mature TOs, TRs, and TDs to DO to confirm affordability. At this stage, TOs, TRs, executability, and affordability come to a balance to create the next iteration of TDs.

With TOs, TRs, TDs, and affordability analysis well developed, a series of reviews are conducted. The first set of reviews are technical interchange meetings, conducted by DE and DT, to obtain stakeholder concurrence with TOs, TRs, and costs. Once stakeholders concur, a final cost model is produced. DE and DT teams then present TOs, TRs, TDs, and affordability analysis to DE and DT executive leadership. This review stage may be iterative. Finally, DE and DT present TOs, TRs, TDs, and affordability analysis to

Program Executives (PEs), Program Directors (PDs), and Program Managers (PMs) in an executive level review. Phase 2 ends when the PEs, PDs, and PMs give concurrence.

3. Phase 3

Phase 3 begins as TOs and TRs are finalized in Phase 2 of the IMTP generation, nominally in February, and lasts roughly six months. Phase 3 is overseen by the DT.

Phase 3 shifts focus from individual test designs to the larger test program. This includes allocation of resources by DT, refined target solutions by TC, and updated cost information by DO. Meeting the requirements for individual tests and the larger test program, while working within programmatic, logistical, and financial constraints, is an iterative process done collaboratively by DE, DT, TC, and DO. Technical interchange meetings are held throughout the iterative Phase 3 stages. These meetings obtain stakeholder concurrence before TOs, TRs, TDs, and cost analyses are finalized.

With individual test events as well as the larger test program assessed, proposed final TOs, TRs, TDs, schedules, and costs are presented at a series of reviews. Each of the reviews is iterative, providing the opportunity for revisions until concurrence is reached. The first of such reviews is with DT and DE leadership. The next review is the executive level review, at which DT presents results to PEs, PDs, and PMs. The final review is the Executive Panel review. Once concurrence from the Executive Panel is received, the IMTP Program Objective Memorandum (POM) Decisions Memorandum is signed by IMTP signatories, a new test baseline is established, and Phase 3 concludes. The test baseline agreed upon in Phase 3 is documented in version 0 of the IMTP document.

4. Phase 4

Phase 4 begins in the middle of Phase 3, nominally in May, and lasts roughly 10 months (roughly six months after Phase 3 ends). Phase 4 is overseen by the DT.

Phase 4 updates the test baseline in response to Office of the Secretary of Defense (OSD) resource management decisions and Program Change Board changes. This requires changes to TOs, TRs, TDs, targeting, schedules, and cost estimates. To accomplish this, the following major activities occur, often simultaneously:

- DE updates TOs and TRs.
- DT synchronizes changes, updating TDs to account for changes to TOs, TRs, targets, countermeasures, and cost estimates.
- TC adjusts target solutions using updated target parameters.
- DO updates cost estimates using updated TDs and new affordability limits.

Once the iterative procedure has concluded, with stakeholders in agreement, TOs, TRs, TDs, schedules, and costs are presented to leadership for review. That review process proceeds first through DE and DT approval, followed by three separate executive reviews. If reviewers request changes, the process returns to the prior iterative and often simultaneous steps. The test baseline agreed upon in Phase 4 is documented in version 1 of the IMTP document.

B. Planning and Execution of Individual Flight Tests

The detailed FT planning required to execute individual FTs begins roughly two years out from the BETD, with analysis activities continuing for a couple months after test execution. Major activities during that time include planning, risk analysis, training, asset reservation or acquisition, execution of the test, and post-test analysis. The process itself includes four phases.[11] These four phases, which should not be confused with the four phases for generating the IMTP, as well as a timeline and brief description of the process are shown in Figure 2-2.

The Mission Director (MD) oversees the detailed FT planning and execution process. The MD may be appointed from any MDA organization or participating element, yet follows MDA's processes regardless of their own origins. Chosen by a DT-led selection board up to 30 months before the BETD, the MD leads the Integrated Event Test Team (IETT), which forms the primary workforce during the detailed planning and execution process. The IETT consists solely of government personnel, and each participating organization or Element nominates an IETT member.

BE: Best Estimated Test Date

Phase	Primary Activities
1	**Plan Test Event.** Test plans include test scenario design, system configurations, required equipment (both HW and SW), asset requests (including ranges and data collection sources), analysis plans, and safety. Assess schedule, mission executability, risk, and cost.
2	**Prepare for Test Event.** Perform a mission analysis review and mission readiness review.
3	**Execute Test Event.** Deploy assets and systems under test. Confirm proper communications and architecture integration. Perform test team rehearsal and readiness review. Execute test. As necessary, assess failures with failure response team.
4	**Analyze Test Data.** Assess if test objectives were met and capture lessons learned.

Figure 2-2. Timeline and Primary Activities for Detailed Test Planning Phases Leading Up to Test Execution.[11]

1. Phase 1

Detailed FT planning begins with the appointed MD reviewing existing test material, such as target details, feasibility and safety assessments, TOs, available funding, potential data collection assets, schedule, and risk areas. The MD accepts management responsibility for the test once the test is confirmed executable within the given cost and schedule constraints. This process is referred to as the Transition Review, the conclusion of which begins Phase 1.

Phase 1 makes up the greatest portion of detailed pretest planning, taking place in the 7 to 24 months before the best estimated test date. Phase 1 is dedicated to test planning, as IETT members prepare the detailed test scenario and assess mission executability, risk, and cost. Lessons learned and BMDS Discrepancy Reports from previous tests are reviewed for applicability to the current FT. Test plans include test scenario design, system configurations, required hardware and software, asset requests (including ranges and data collection sources), analysis plans, and safety.

Phase 1 products are reviewed midway through the phase at the Test and Readiness Review, nearing the end of the phase at the Mission Planning Review, and at the conclusion of the phase at the Executive Mission Planning Review. The Test and Readiness review addresses TOs, plans, and risks. The Mission Planning Review considers all Phase 1 products as well as risk posture. The Executive Mission Planning Review approves TOs, test configurations, risks, and schedule.

2. Phase 2

Phase 2 begins detailed planning for mission execution, and takes place two to seven months before the best estimated test date. This detailed planning includes pre-deployment plans and in-depth risk analysis. The start of Phase 2 also signals an Operational Test and Readiness Review (OTRR), which is the first formal tasking for operational testers in the detailed planning of a single specific test event (i.e., mission-director-led activities). At OTRR 1, operational testers identify constraints to TOs and coordinate corrective actions.

Throughout Phase 2, efforts are made to assess and mitigate risk. Plans for logistics, analysis, safety, and communications are documented. Digital performance analyses and system pre-mission tests verify mission timelines and system interoperability. Frequency assignments are deconflicted, and sufficient margins for radio frequency communication are confirmed using link margin analysis.

Phase 2 concludes with a series of reviews. The Mission Analysis Review assesses all pre-mission analyses, the Test Resources Readiness Review assesses the readiness of test resources, and the Mission Readiness Review considers the mission risk posture and composite risk assessment. Finally, a summary of the Mission Readiness Review material is presented at the Executive Mission Readiness Review before final approval is provided for transitioning to Phase 3.

3. Phase 3

Phase 3 includes test execution and the two months leading up to it, during which all preparation activities are finalized.

An Integrated Master Schedule guides asset deployment. Generally, assets are deployed in Phase 3, though long-lead assets may have been deployed near the end of Phase 2. Systems under test and targets are deployed following satisfactory readiness reviews. Plans for handling and analyzing data are finalized. Frequency assignment activities are completed and the communications architecture is set. The Mission Execution Team rehearses in preparation for test execution. These pre-mission training activities take place both prior to deployment to the test site and after, even including a dry run immediately prior to the test event.

A series of reviews are carried out before test. The Mission Execution Review evaluates readiness, before the Executive Mission Execution Review accepts risk and grants approval for test execution. However, the Executive Mission Execution Review may take place weeks before the test is carried out, and two additional readiness reviews, the Launch Readiness Review and Launch Decision Meeting take place a day and then only hours, respectively, before execution. Following execution, the "as-run" test configuration is recorded. Phase 3 concludes with the MD releasing the Mission Execution Team and authorizing a break of configuration (test assets no longer restricted to test positions and setup). The post-event process is expanded in the event of a test failure, with the test configuration remaining locked and a review carried out by a failure response team.

4. Phase 4

Phase 4 takes place in the two months following test execution. Phase 4 begins with the breaking of test configuration, as test sites are cleared, assets are recovered, and post-flight environmental actions are performed. As an exception, tests that fail will permit the start of Phase 4 analysis activities while assets remain locked in position.

Analysis and reporting activities occur in a quick-look increment and a detailed increment. The quick-look portion occurs in the two weeks following test, as initial insights to test results are presented at the Quick-Look Briefing followed by the Executive Quick-Look Briefing. These initial insights also suggest areas of focus for the detailed increment. The detailed-look portion begins after the Quick-Look Briefing, lasting roughly two months as the test community reviews the test data. Similar to the quick-look briefings, a Mission Data Review and Executive Mission Data Review are held. The Executive Mission Data Review formally marks the conclusion of the test event.

There are two other items of note: First, in the event of a test failure, the process also includes a Failure Reporting, Analysis, and Corrective Action System Review to determine if additional actions are required. Second, all test participants, from all test phases, will collect lessons learned to be shared at the Lessons Learned Review a month after test. Lessons learned is its own process, external to test, with an accompanying database and annual DT brief. It was lessons learned, like these, that were reviewed in Phase 1.

C. Chapter 2 Findings

Finding 1: The heavily iterative nature of generating a single IMTP requires substantial input from multiple stakeholders.

IMTP generation begins with new or updated TOs and TRs, then successively adds TDs, target details, and costs, before finally synchronizing with the larger test program. At each iteration, stakeholders are asked to review. This helps ensure stakeholder needs continue to be met as planning evolves. Later iterations, specifically those found at the end of phases, include formal agreement toward approval of the test baseline. Changes to test plans at later stages are more cumbersome to make than changes from earlier iterations.

However, a point often shared during interviews was that stakeholders would occasionally skip MDA-intended feedback opportunities, instead providing such input at later stages. This late stakeholder feedback disrupts the iterative process, rendering earlier iterations moot. MDA is then forced to make last-minute adjustments in an attempt to make up the difference.

Inspection of MDA's process for IMTP generation shows that 2/3 of the IMTP generation steps require stakeholder coordination. The breakdown of these stakeholder steps across each formal phase of the process is shown in Table 2-1. The counting shown in Table 2-1 is conservative, as steps that "incorporate" stakeholder feedback but do not "request" actions during that step from stakeholders do not influence the count.

Table 2-1. Steps at each phase of IMTP generation that request stakeholder coordination.[10]

Phase	Steps	Steps requesting stakeholder coordination
1	5	4
2	7	5
3	11	7
4	11	7
Total	**34**	**23**

Finding 2: Operational testers, external to MDA, are engaged late in the detailed FT planning process.

For each FT, 24 months prior to the best estimated test date, detailed planning begins. An MD leads the test preparation, with a support force of an IETT. Members from the IETT are assigned various tasks to perform and products to produce.

As members of the IETT, operational testers are invited to IETT meetings. However, operational testers are not explicitly tasked until seven months prior to the best estimated test date as shown in Figure 2-3. That first explicit engagement is the OTRR, an event intended for operational test agency leadership and with no immediate product for MDA. In fact, the event may be viewed with such little significance by MDA that MDA neglected (likely due to oversight) to place it in the Phase 2 Tasks and Products table in the governing documentation[11] despite the event being on the Phase 2 schedule.

To be clear, operational testers are involved in the test planning process prior to the first Operational Test and Readiness Review, including overall flight test planning (like that used to generate the IMTP). The OTRR represents their first formal tasking for a specific flight test once the test is under the purview of an MD, as captured in the MDA

Directive 3000.10 task lists. Operational testers are also involved in understanding the results of tests and their implications for the content and number of future tests.

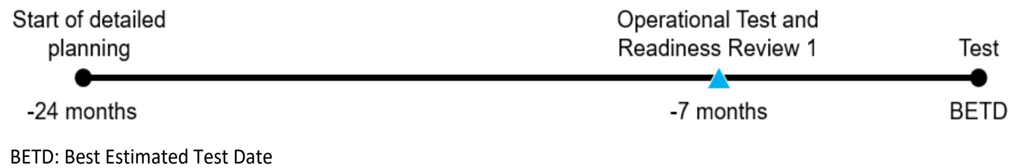


Figure 2-3. Timeline Showing Operational Testers' First Explicit Tasking, the OTRR 1, in the Two Years Leading up to Test Execution.[11]

D. Chapter 2 Recommendations

The following two recommendations, related to MDA's FT planning process, pair with the above two findings.

Recommendation 1: IDA recommends that MDA review the IMTP generation process and determine if stakeholder coordination is truly required for each of the 23 process steps currently listed.

For steps in which stakeholder coordination cannot be avoided, MDA should identify which steps are particularly costly or disruptive when stakeholders do not provide input as intended. MDA should attempt to revise its operations to mitigate such anticipated yet especially significant disruptions. MDA's expectations of stakeholders should be relative to their size of support (including both internal and external IMTP contributors).

MDA may also find success in educating stakeholders on the importance of engagement and the consequences of not providing input at each step. Stakeholders should respect MDA's process and understand that late input may not be accepted. Stakeholders may be reluctant to oblige when they are asked for feedback too frequently. However, a more spaced-out feedback schedule may improve the timeliness of stakeholder engagement at each step.

Recommendation 2: IDA recommends MDA use operational testers during the first year of detailed FT planning, beginning with the Kick-off Planning Meeting.

The first formal task for operational testers occurs 17 months into the 24 months before test - the OTRR 1. At that time, operational testers review requirements, test plans, and issues. Prior to OTRR 1, operational testers have seemingly at-will participation in the IETT. However, operational testers should instead support test planning activities as early as feasible and any early ad-hoc support should be formalized. Additionally, to take advantage of early engagement, operational testers should look for data collection opportunities as they support the above tasks.

The next iteration of MDA Directive 3000.10 should include OTRR 1, in its Phase 2 Tasks and Products (currently Table 5), likely with the operational test agency as the performer, and the Operational Test Mission Director as the IETT point of contact. The three OTRRs should produce products to be shared with MDA.

(This page is intentionally blank.)

3. Assessment of MDA's Flight Test Schedule Revisions and Instabilities

This chapter describes IDA's assessment of MDA's FT schedule revisions and instabilities. The study team assessed the extent to which MDA has met its goals for past annual BMDS FT plans, and to identify obstacles and challenges in meeting these goals, including developmental delays, system readiness, and range and target availability. The methodology used and its limitations are briefly described before the discussion of the examination of FT plans and execution, schedule changes, reasons for changes, and achievement of goals.

A. Assessment Methodology

To address FT schedule revisions and instabilities, the team's methodology included examining the FT schedules as laid out in the IMTPs and related Smart Books[12] as well as the annual DoD fiscal year (FY) President's Budget (PB) Submissions documented in Missile Defense Agency Justification Books. Generally, each FT in the IMTP and in the budget submissions has a specific quarter and year for planned execution; IDA analyzed changes from year to year. Across the set of IMTPs, the frequency of changes could also be evaluated. Later IMTPs and Smart Books provide a rationale for FT schedule changes. The study team established a taxonomy to determine if there were underlying trends or process issues that could be identified from the reasons listed.

There were some limitations to this approach:

- The IMTP represents a snapshot of the test baseline at specific points in time, rather than an in-depth examination of execution year test scheduling. Because the FT dates in the IMTP are specified by quarter and year, the granularity of test date changes is limited to three-month increments.
- There are gaps in the data. For example, the PB FY19 Justification Book says that the IMTP is at a higher classification and does not have any FT schedule data, while PB FY22 does not include any out years. Starting with Version 20.1 in 2019, Critical Engagement Conditions (CECs) and Empirical Measurement Events (EMEs), which gauge the collection of data from flight and ground tests,

[12] The IMTP Smart Book is "an FOUO summary of the IMTP with unclassified portions of the database. A quick reference guide into the test program primarily distributed within MDA and external stakeholders."

were no longer listed in the published test plans. Many of the rationales for schedule changes are not recorded in the IMTP or Smart Books before Version 17.1 (22 April 2016).

- The analysis had to have a starting point. For detailed changes in schedule, the study team started with V14.1 (31 March 2014) because prior versions of the IMTP Smart Books did not include tabulated data like range, target, and supported increment. The supported increments also changed names around that time (V14.2).
- Although IDA is considering only FTs and not exercises, some exercises like Formidable Shield include intercept and target-only FT events. Some FTs are described as tracking "exercises." For the purposes of accounting, all are included.

Subject to these limitations, the team built an overall picture of the FT schedule as it has evolved over the years.

B. Flight Test Plans and Execution

This section details trends observed in the FT data from both the IMTP and PB submissions. The records for IMTP Smart Books begin with V10.2, dated August 25, 2010, and the final Smart Book the team examined was V22.1, dated May 27, 2021. On average, there were 62 FT records per IMTP, with approximately 1,200 total records (note that there can be up to two IMTPs per year). Early IMTPs (through V12.2) have tests listed over a period of 12 years, while more recent ones list tests over seven years. There are 341 unique FT names, with 91 having completion dates scheduled beyond May 2021. Each test appears on average in three IMTPs. Some FTs have been renamed or combined over the years.

For the PB data submissions, the records start with the FY2011 PB Justification Book (dated February 2010) and run through the FY2022 PB Justification Book (dated May 2021). For the PB data, there were on average 59 FT records per PB with approximately 650 total records. The PBs have tests listed five years into the future. There are 257 unique FT names in the PBs, with 66 having completion dates scheduled beyond May 2021. Each test appears on average in 2.4 PBs.

1. Flight Test Planning

The study team looked at the summary tables provided in each IMTP Smart Book since V10.2. In addition to the number of tests, these summary tables provide counts of salvo/multi-intercept missions as well as a count of interceptors and targets. Figure 3-1 shows the totals by IMTP for FTs, Salvo Intercepts, Interceptors, and Targets from the IMTP Smart Book summaries. The annual average across IMTPs is 20 salvo/multi-intercept missions, 92 interceptors, and 82 targets for each IMTP. There is a significant increase in FTs, interceptors, and targets planned in V22.1; the reasons for this will be discussed later in this chapter.

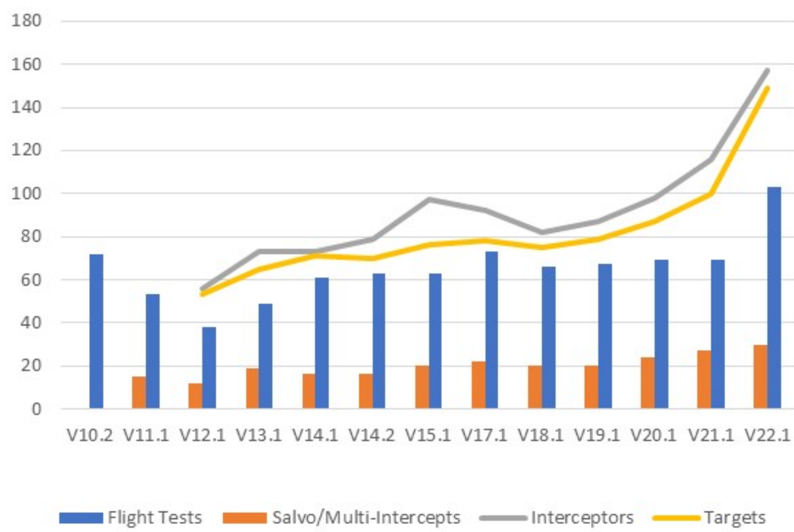


Figure 3-1. IMTP Summary Totals

The study team looked at the commonality between the FTs in the Smart Books and those in the PB. In addition to being two separate sets of data, the comparison illustrates differences between planning and budgeting. Figure 3-2 shows the count of tests for each, with the PB fiscal year aligned to the corresponding IMTP. There is reasonable agreement between the two. Since V14.1 (March 2014), the average difference has been about three percent excluding V19.1 (no data for the PB) and V22.1 (no out-year tests reported in PB). Part of the discrepancy appears to be due to the PB having a more limited time horizon as well as listing completed tests where post-test activities are still being budgeted. Starting with detailed data from V14.1 and onward, tests common to both the PB and IMTP have an average test date within two years, while the tests that are not common to both the PB and IMTP have an average test date about 4.5 years out.

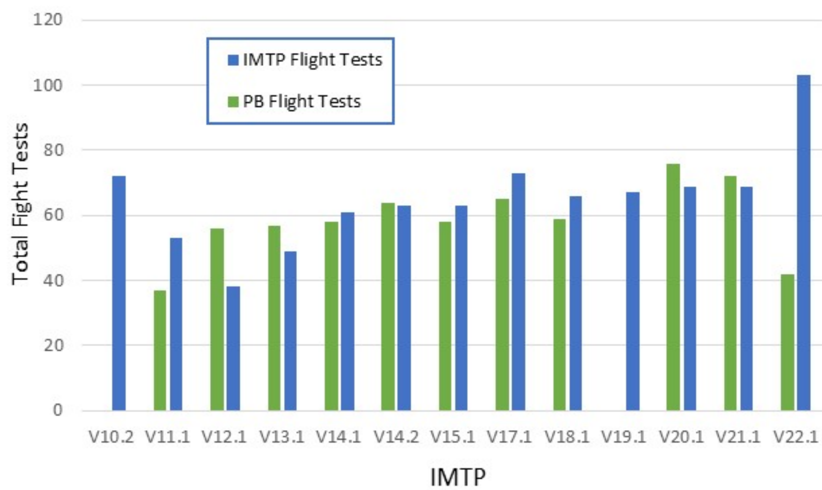
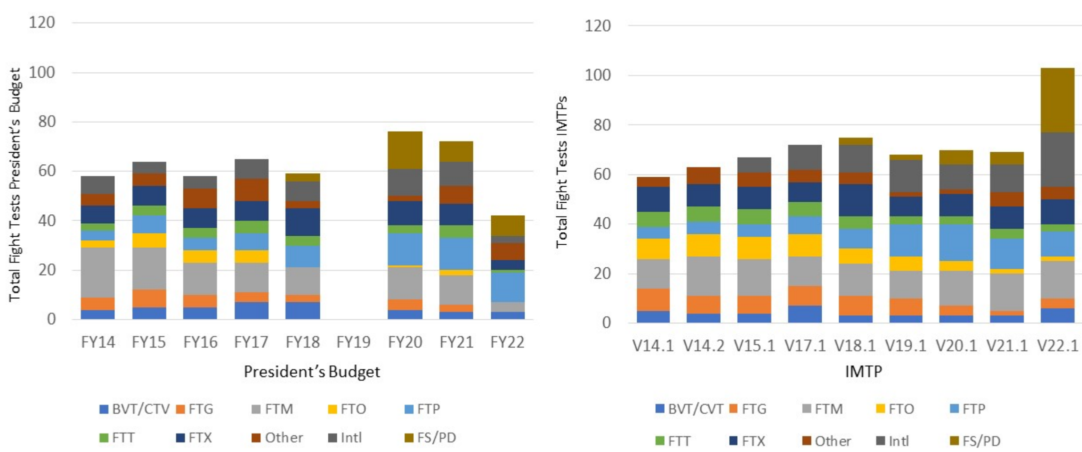


Figure 3-2. Comparison of Flight Test Counts

Figure 3-3 shows the distribution of FTs by type for the PB documentation (on the left) and the IMTP since V14.1 (on the right). The FY22 PB did not have any FTs listed for the out years, including some international tests and FT events from the Formidable Shield and Pacific Dragon exercises, which, as seen in the right chart in Figure 3-3, comprise the large increase in the number of FTs in the IMTP over the annual average. This graphic also shows the differences between planning and budgeting; for example, for the GMD tests labeled Flight Test GMD (FTG), there are more in the IMTP than in the PB. Operational FTs also appear to occupy a larger fraction of tests in the IMTPs than in the PB across the years.



Key: FTM = Flight Test Aegis, FTP = Flight Test Patriot, FTX = Flight Test Other, FTG = Flight Test GMD, FTT = Flight Test THAAD, FTO = Flight Test Operational, BVT/CTV = Booster Vehicle Test / Controlled Test Vehicle, FS/PD = Formidable Shield/Pacific Dragon, Int'l = International cooperation

Figure 3-3. Comparison of Flight Tests by Type, PB (left) and IMTP (right)

IDA also used IMTP data to examine FTs by BMDS Capability Increment. Figure 3-4 shows the number of FTs in the IMTP by the BMDS increment supported. In general, the number of tests for early increments gradually decreases while tests for later increments are added, consistent with planned incremental capability delivery. Increment 5 will be examined in detail in a later section. This graph starts at V14.2 because the BMDS increments had different names in prior IMTPs.

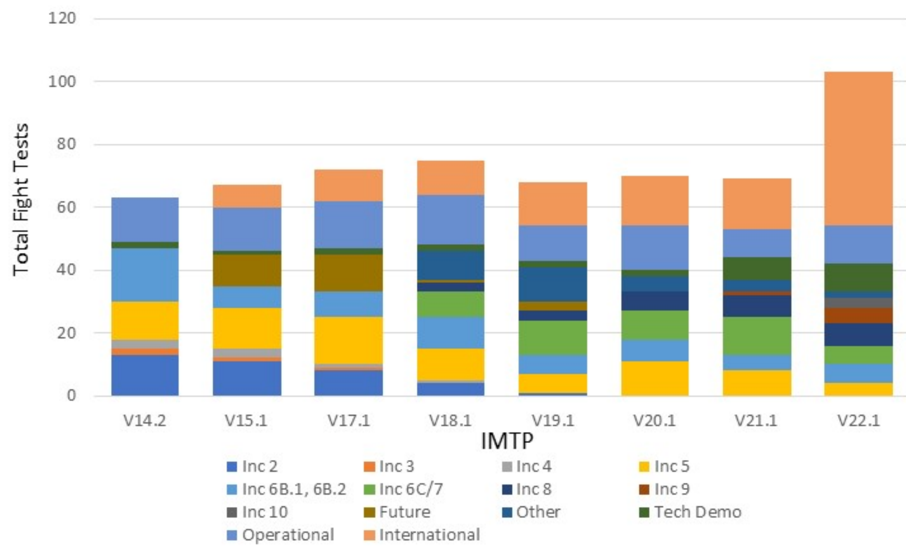


Figure 3-4. Flight Tests by BMDS Capability Increment

There is an increase in the average number of planned FTs per year, but the number of flight tests supporting specific BMDS increments stays roughly constant. MDA's V22.1 IMTP has a large increase in FTs over the annual average. This is mostly due to international tests and FT events in exercises. There are increases in the numbers of planned salvo intercepts, interceptors, and targets used for flight testing. The trend of more tests with more complexity will be explored in more detail in a later section.

2. Flight Test Execution

IDA examined the GAO result stating that, between FY10 and FY19, MDA conducted 37 percent of its baseline fiscal year testing as originally planned. Figure 3-5 shows the GAO result for FTs conducted and FTs delayed, deleted, or merged, as derived from Figure 2 in GAO 20-432.[13] For this graphic depiction of the GAO analysis, no-tests

[13] Missile Defense: Assessment of Testing Approach Needed as Delays and Changes Persist GAO-20-432 Published: Jul 23, 2020.

were counted as conducted since resources and planning efforts were dedicated to them. With no-tests included, the percentage of tests conducted is 39 percent.

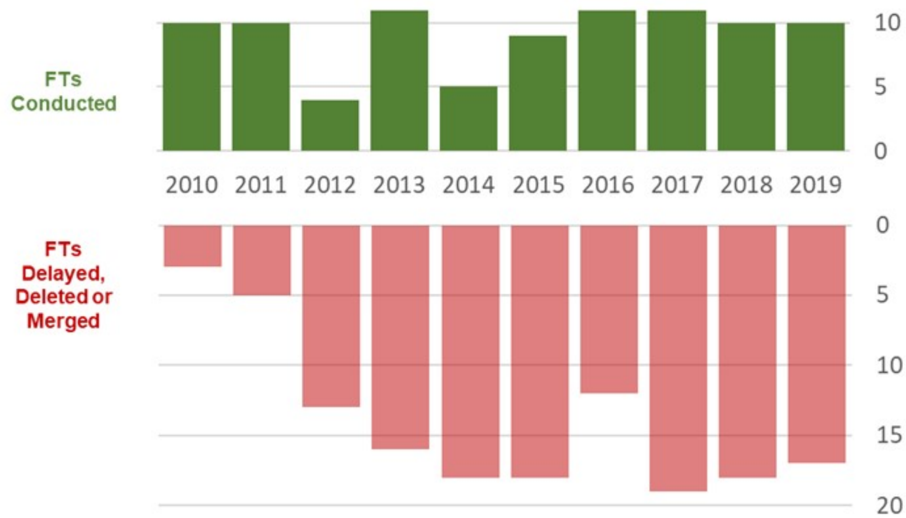


Figure 3-5. GAO Analysis

Including the no-tests, IDA analysis also shows 39 percent, with what is assumed to be similar methodology, and 37 percent if the no-tests are not included. All Patriot and Israeli system tests were excluded. Across all IMTPs, if a FT was ever planned for a particular fiscal year it was counted. There are slight discrepancies between IDA's and GAO's analyses in what appear to be counting of exercise events that do not appear in the IMTPs but are executed, as well as the merging and changing of test names. Figure 3-6 shows the IDA analysis.

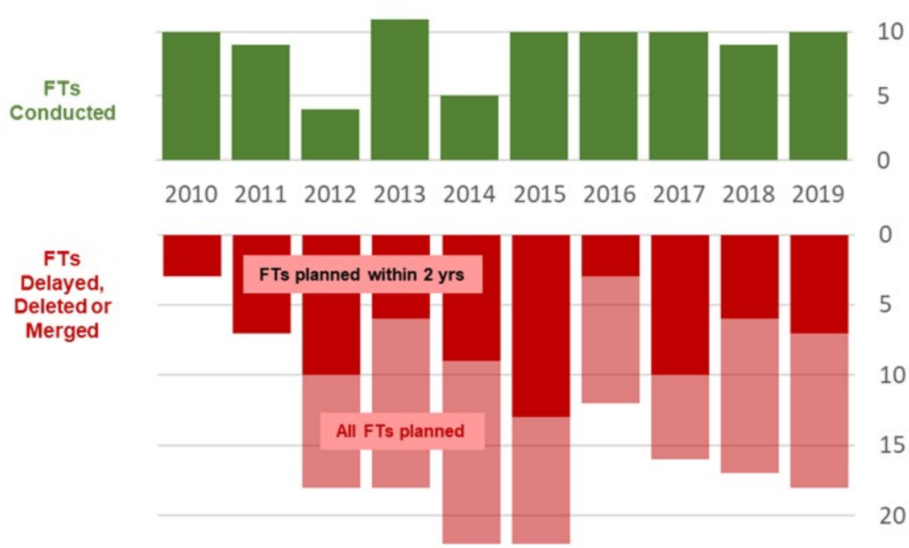


Figure 3-6. IDA Analysis of Flight Test Execution

As mentioned previously, early IMTPs spanned as many as 12 years while more recent ones cover about seven years. A test that may have been moved multiple times over the years is counted as a change in every fiscal year it was scheduled until it is executed, merged or deleted. The methodology captures both test accomplishment and schedule changes in a single number. The method does not include the reasoning for or resulting consequences of a schedule change. Changes in FT scheduling may be intentional or unintentional, and may be due to factors outside MDA's direct control.

Detailed FT planning begins about two years before the test date. Thus, IDA also looked at test execution considering only tests that have been planned within two years. Under that assumption, the test execution increases to 54 percent of planned tests, as indicated by the darker shade of delayed, deleted, or merged tests in Figure 3-6.

The team also extended the analysis past the GAO cut-off date of FY19. Figure 3-7 shows the same chart with the addition of FY20 and FY21 results. COVID disrupted the planned FT schedule in FY20. The measure of FT execution to FTs planned drops to 35 percent considering all tests over all IMTPs and remains 54 percent when considering only tests planned within two years.

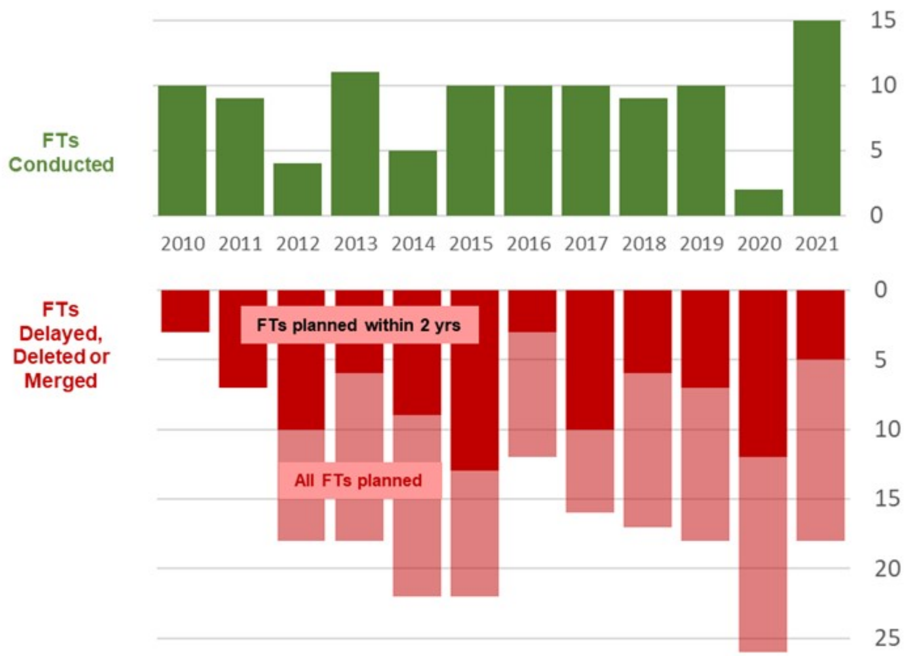


Figure 3-7. Flight Test Execution through FY21

Over the decade FY10-19, MDA executed an average of about 8.8 planned FTs annually, and this number does not change significantly (8.75) when including FY20-21. The median number is 10 tests per year in both cases. A maximum of 15 FTs were executed in FY21 not including Patriot and Israeli system tests. The average number of tests planned in each year across IMTPs is 22.5, but this number reduces to 16.3 if only planning two years beforehand is considered. This difference can be understood as the difference in scope between long-term planning and budgeting.

Four of the events executed in FY21 were Formidable Shield FT events, demonstrating the ability to conduct additional tests beyond the average number listed with international partners. This also seems to indicate that the large number of planned FT events in exercises in V22.1 may be executable over the historic average number of BMDs increment flight tests.

C. Flight Test Schedule Changes

The team investigated FT schedule stability by looking across time as well as annually. Using detailed schedule data from IMTP V14.1, approximately 645 records, FT date changes were examined. There were 224 FTs in this dataset, 200 of them had a quarter and year specified for the test date, and 82 of these have completion dates beyond May 2021. There were 88 tests with changes in completion date, or 44 percent. The median number of days from the original to the new test date is 548. Table 3-1 shows the breakdown by select BMDS elements. Aegis, GMD, and THAAD have a high proportion of FTs with delays compared with some of the other programs shown in the table. As of V22.1, no FT events in exercises have been delayed.

Table 3-1. Flight Tests by Type from IMTP V14.1 to V22.1

			Delay (days)		
	Total Tests	Tests with Delays	Min	Median	Max
Aegis (FTM)	28	22	91	457	3,653
GMD (FTG)	9	8	365	914	3,470
THAAD (FTT)	9	7	90	365	1,735
Patriot (FTP)	28	16	0	92	731
Other (FTX)	27	13	0	0	1,945
Operational (FTO)	10	4	0	0	3,287
Exercises FTs (FS/PD)	36	0	0	0	0
Vehicle (BVT/CTV)	19	8	0	0	823

From the FY11 PB to present, about 43 percent of tests have changes in completion date, and over the period as shown from FY14, the PB data shows about 40 percent. This appears to be consistent with the IMTP.

Figure 3-8 shows the FT date changes for Aegis, GMD, and THAAD graphically. If there were no schedule changes, a FT would appear as a horizontal line in these graphs.

Tests that experience delays are represented by lines with upward turns. Multiple turns for a test indicate multiple changes over the years.

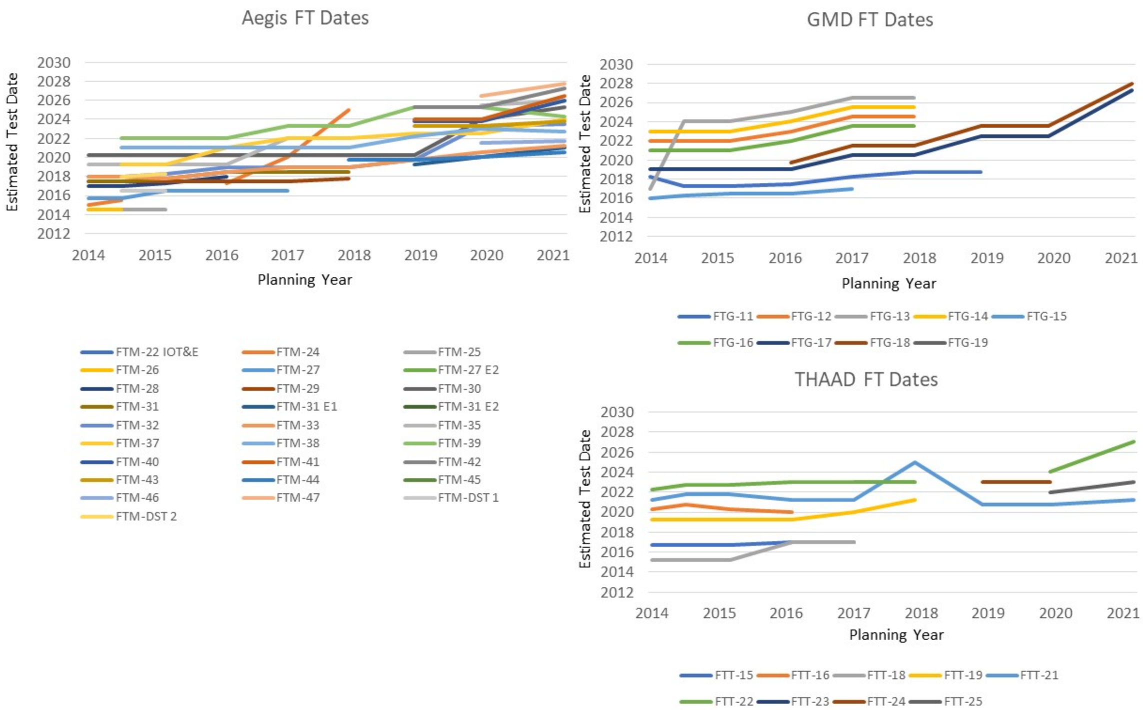


Figure 3-8. Flight Test Changes for Aegis, GMD, THAAD

In addition to tracking FTs across the IMTPs, IDA also looked at what percentage of each IMTP is changing every year. There is an average of 56.5 FTs in each IMTP and an average of 20.7 changes in test date (excluding new tests appearing as a change). Therefore, in any given IMTP, an average of 36 percent of the tests are being rescheduled, deleted, or merged. The team also examined the rolling average of FTs with changes each year; Figure 3-9 shows the result using a three-year period. The rolling average was relatively stable until the large number of changes caused by COVID in V22.1. Excluding V22.1, the average number of tests with schedule changes and the moving average were fairly consistent at around 33 percent per year. Within the limitation of the IMTP being a snapshot of the test baseline at a particular point in time, the fraction of the IMTP that changes each year appears to be consistent.

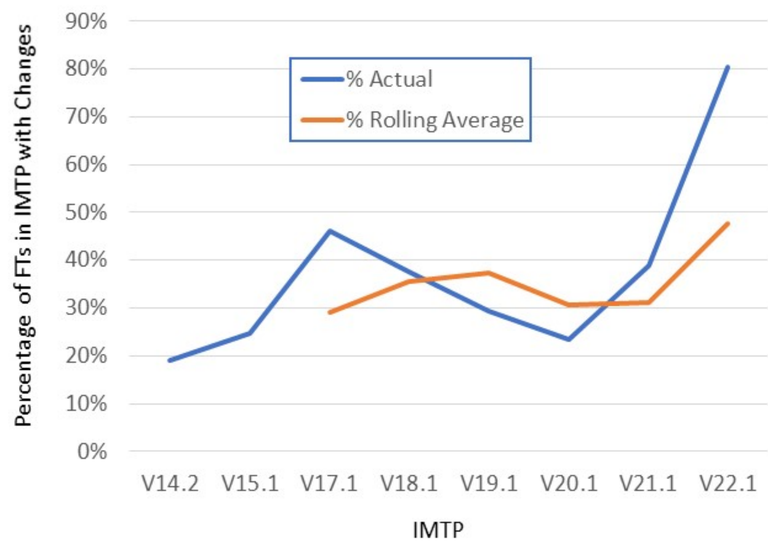


Figure 3-9. Rolling and Annual Average of Tests with Changes

D. Increasing Test Quantity and Complexity

The study team explored one additional aspect of planning and execution: increases in the number of FTs and their complexity. During many interviews, the general impression was conveyed that the scope of testing has been increasing at a rate greater than resources are being made available. Some of the points made were (1) systems under test and the scenarios designed to test them were becoming more complex each year, (2) funding and the workforce were not scaling at the same rate, and (3) aging test ranges, supporting infrastructure, database systems, and software tools were all contributing to increased challenges.

The IDA team investigated these assertions using available data. From Figure 3-1, the number of planned FTs is increasing each year by about 3.8 percent of the annual average. Across the IMTP totals for each year, there are increases over the annual average of planned FTs, planned salvo tests (6.4 percent), the number of interceptors (11 percent), and the number of targets (11.7 percent). Because IMTPs have tests planned between 7 and 12 years, the study team also examined the average annual FTs planned for the immediate three years following IMTP publication. Figure 3-10 shows the results, which indicate 5.4 percent growth in planned FTs over the annual average. The growth in FTs conducted, based on the green bars in Figure 3-7, is 5.2 percent over the annual average due in large part to the increase in flight tests in FY21.

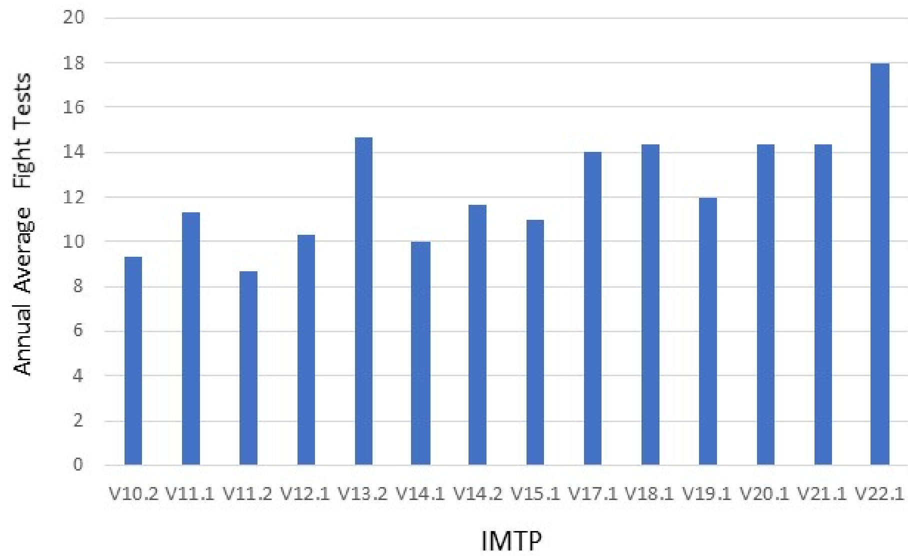


Figure 3-10. Average Annual Planned Flight Tests, within 3 years of IMTP

MDA total test-related spending is also increasing.[14] IDA found an average $32 million increase per year, or an increase of 3.7 percent of the annual average each year. Isolating FT costs proved to be difficult, so the team took the PB numbers for all test-related spending that could be identified. The largest category was “BMD Test” but there were related lines in other budget elements including test-related categories for C2BMC, Aegis, Sensors, and others. Figure 3-11 shows the breakdown of these budget contributions.

[14] IDA was not provided data for just flight test spending.

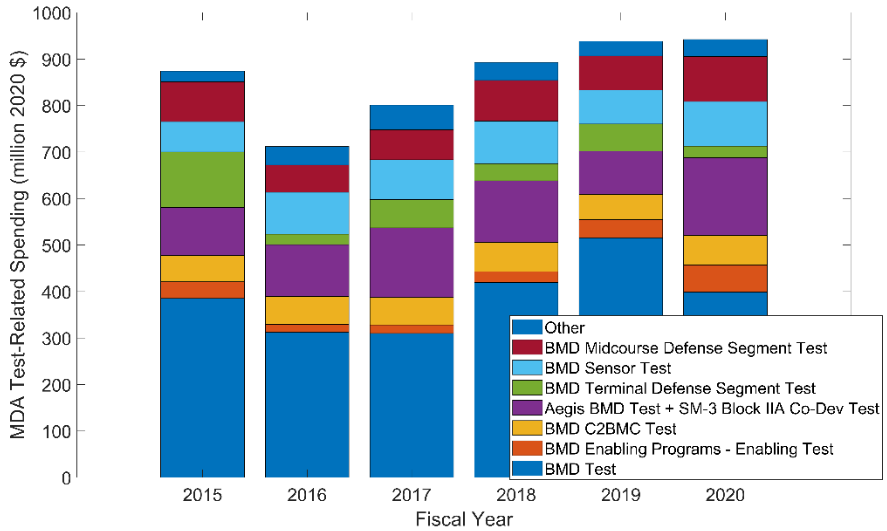


Figure 3-11. MDA Total Test-Related Spending

Although the number of tests and test-related spending are increasing at roughly the same percentage, by some measures FTs are becoming more complex. Across IMTPs, the study team observed increases over the annual average in: tests that included both air and ballistic targets (7 percent); tests with two or more different types of interceptors (14 percent), and tests where the target and shooter are in different ranges (16 percent). Figure 3-12 shows these tests both in total number and as a percentage of tests in the IMTP.

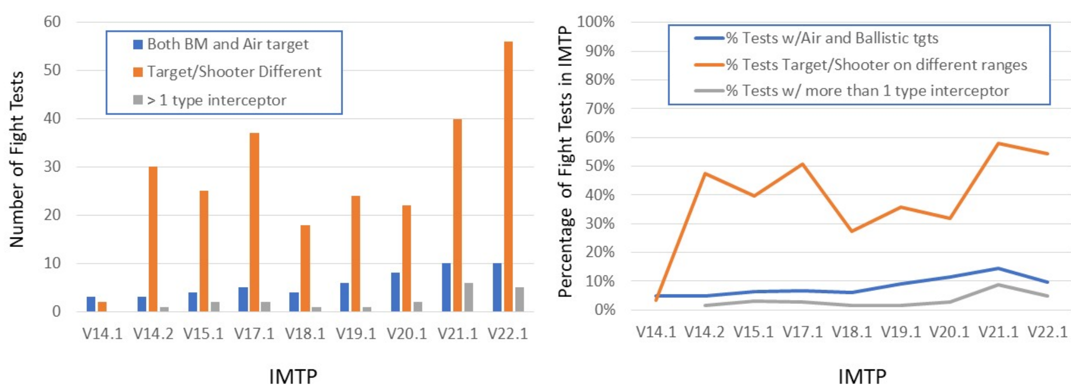


Figure 3-12. Number and Percentage of FTs in IMTP and Select Measures of Complexity

There are 53 total records that show both an air and ballistic missile target. About a quarter of these tests are for Aegis (FTM), a quarter for Patriot (FTP), a quarter for Operational (FTO), and the remaining quarter for others that include sensors (FTX), exercise events (FS/PD), and international partners. Of the 20 test records with more than a single type of interceptor, 18 of them involve combinations of Standard Missiles (SM-2, SM-3 Block IA/IB, SM-3 Block IIA, and SM-6). Table 3-2 shows the most common combinations in the FT records for shooter launch range and target launch range.

Table 3-2. Most Common Shooter/Launch Range Combinations

Shooter Launch Range	Target Launch Range	Total
PMRF	PMRF	146
WSMR	WSMR	83
PMRF	Hickam	54
VAFB	RTS	38
Kodiak	Hickam	22
BOA	Hebrides	20
RTS	RTS	20
RTS	Hickam	19
WSMR	Ft Wingate	15
Wake	Hickam	14
Wallops	Wallops	13
Pt Mugu	Pt Mugu	11

BOA - Broad Ocean Area, PMRF - Pacific Missile Range Facility, RTS - Reagan Test Site, VAFB - Vandenberg Space Force Base, WSMR - White Sands Missile Range

1. Increment 5 Test Complexity

There is additional information in the IMTPs on test resources and system participation. However, most of this information is in graphical stoplight chart form and not readily convertible for analysis. The study team looked closely at a single BMDS increment's configuration sheets from the IMTPs (Increment 5; European Phased Adaptive Approach Phase 3 and Robust IRBM Defense). Detailed information from IMTP V15.1 through V20.1 was examined for FT complexity.

For the years in question, data from the Smart Books show that the number of FTs supporting Increment 5 fluctuates over time and actually decreases overall. It also shows that the number of interceptors per FT and the number of targets per FT are staying roughly constant, with a 0 percent increase over the annual average. The Smart Book data show that the number of FTs where the target launch range and interceptor launch range are different are increasing at 14 percent over the annual average. These trends are graphically presented on the left of Figure 3-13.

The IMTP data for Systems Under Test (SUT), Developmental Systems Under Test (DSUT), Associated Systems, Test Resources, and Operational Assets are shown on the right side of Figure 3-13. The number of Test Resources per FT is increasing at 4 percent over the annual average, while the number of SUTs participating per FT is increasing at 7 percent. The other categories appear to stay relatively steady until V20.1 where they all show a sudden increase. From the limited IMTP data, the team was able to determine that for FTs supporting Increment 5, the average number of planned participants in each FT is increasing from 16.3 in V15.1 to 25.6 in V20.1, about 12 percent per year.

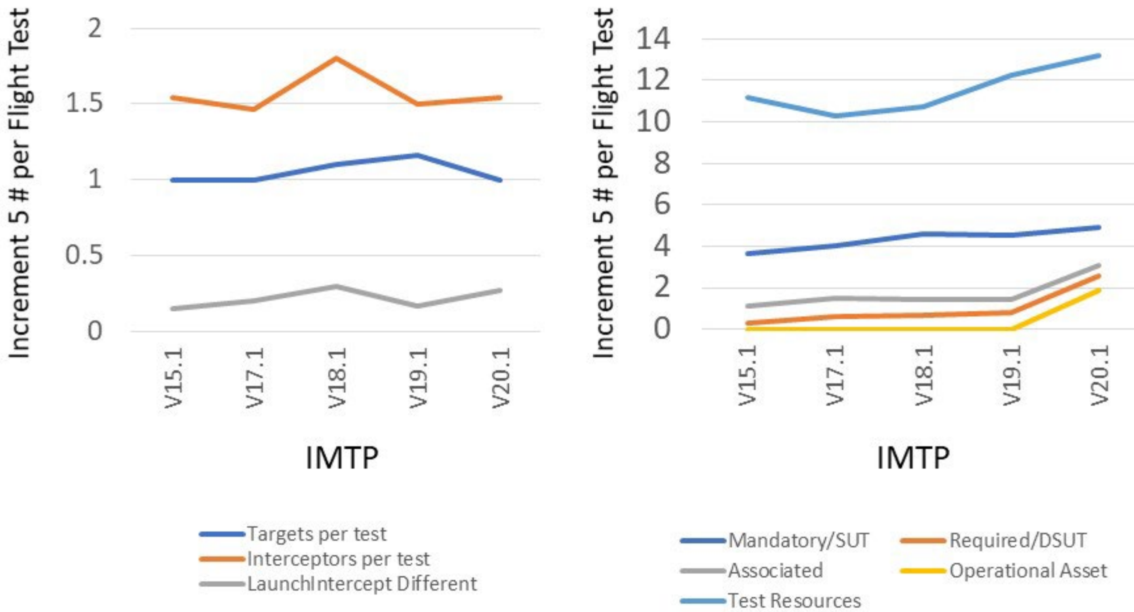
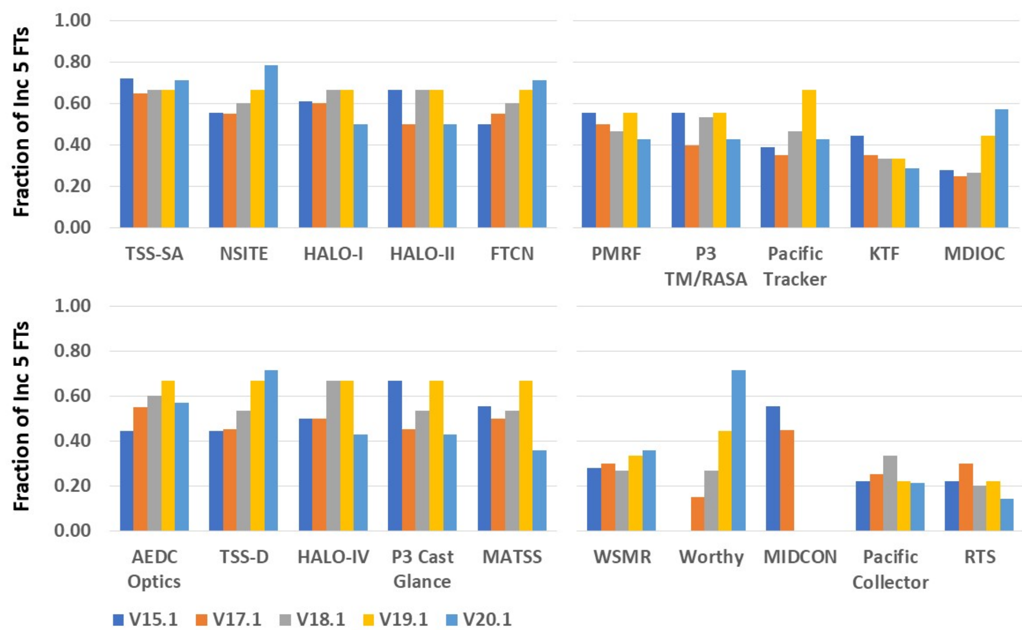


Figure 3-13. Increment 5 Measures of Flight Test Complexity

IDA looked at the test resources listed in the IMTPs. The fraction of Increment 5 FTs that MDA test resources plan to support is shown in Figure 3-14. The Test Support System – Situational Awareness (TSS-SA) is intended to participate in nearly 70 percent of the tests on average across IMTPs, while the Pacific Collector is planned to support about 25 percent of Increment 5 FTs. The participation of some resources, like the Flight Test Communications Network (FTCN) is steadily increasing with each IMTP while the participation of other resources such as the Kauai Test Facility (KTF) is decreasing. The main takeaway is that MDA has data available to investigate test resource dependence and reliance. If there are assets that have availability or reliability issues but are in high use, it may be possible to anticipate FT schedule delays due to test resources.



AEDC - Arnold Engineering Development Center, FTCN - Flight Test Communications Network, HALO - High Altitude Observatory, KTF - Kauai Test Facility, MATSS - Mobile At Sea Sensor System, MDIOC - Missile Defense Integration and Operations Center, MIDCON - Missile Defense Comm and Operations Node, NSITE - Network System Integration and Test Environment, PMRF - Pacific Missile Range Facility, RASA - Remote Area Safety Aircraft, RTS - Reagan Test Site, TSS-D - Test Support System Data, TSS-SA - Test Support System Situational Awareness, WSMR - White Sands Missile Range

Figure 3-14. Top 20 Test Resources Supporting Increment 5

This section examined the possibility that MDA is trying to do more with less resources. From the data examined, it appears that FT numbers and overall test budgets are rising at approximately the same proportion, and by some measures, the FTs being planned are getting more complex.

E. Causes of Test Delays/Cancellations

Beginning with IMTP V17.1, a detailed rationale for FT changes has been included in the IMTP, leading to approximately 350 records of schedule changes through V22.1. The study team looked at these rationales to see if there were common issues and trends. Some of the challenges included the level of detail and attribution when multiple reasons were provided. Some examples include:

- "Accelerated due to COVID-19 impacts and deconfliction of FY 2020 Test events."
- "Delayed due to programmatic reasons."
- "Delayed due to developmental delays in software deliveries to support flight test."

- “Delayed due to a fact of life change.”

Of the 350 records, there were approximately 240 unique rationales. To simplify the analysis, IDA constructed a taxonomy to characterize FT change reasoning. For this taxonomy, the team focused on FT schedule delays, deletions, and modifications and ignored additions, renaming, and accelerated tests. In total, there were 211 records for delayed, deleted, or modified tests.

Major factors contributing to FT delays were identified from the IMTP rationales. Some of the rationale language was descriptive and some is less so. The study team summarized these descriptions using as few words as possible. In some cases, there were two reasons given and they were both given short descriptions and assessed for the main and secondary reason.

These were then categorized in a two-by-two classification matrix: one axis distinguishing intentional and unintentional causes of delay, and the second axis differentiating internal versus external causes. Under this taxonomy, intentional refers to what appear to be deliberate decisions to ensure test cadence and execution of the baseline, while unintentional refers to unplanned and unexpected issues that arise. Internal refers to causes that appear to be under the control of MDA or its program offices, while external refers to factors that are outside MDA’s direct control.

For example, “Deleted all previous Redesigned Kill Vehicle (RKV) related flight tests from IMTP v22.1.” was summarized as “RKV Cancellation” and classified as an external, intentional decision. “Direction of the Army’s Lower Tier Project Office” was summarized as “Program direction” and also classified as an external, intentional decision. “Delays due to COVID-19” became “COVID” and classified as external and unintentional. “Delayed due to developmental delays in software deliveries to support flight test” became “SW programmatic” and was classified as internal and unintentional. Table 3-3 shows this matrix and the high-level short descriptions.

Table 3-3. Taxonomy for Flight Test Schedule Changes

	Unintentional	Intentional
Internal	Developmental Issues HW/SW Availability HW Qualification Programmatic Prior Test Failure SW Programmatic	Added Test[15] Affordability Align Capability Delivery Change of Objectives Configuration Change Prioritization Test Separation Transfer of Objectives
External	Asset Scheduling COVID Range Availability Target Issues	Change Targets FMS Partner Request Program Direction RKV Cancellation

Of the 211 records with change rationales for delay, deletion, or modification, 49 did not include a complete date (missing quarter) to determine the extent of the change. This led to 162 records with enough data for quantitative assessments of the time between the change action and the FT date. The breakdown of these is shown in Figure 3-15.[16] About two-thirds of FT delays, deletions and modifications appear to be internal to MDA, divided equally between intentional and unintentional causes. The other third is due to factors external to the agency. About half of all delays, deletions, and modifications appear to be intentional.

[15] For this taxonomy, added test refers to a delay, deletion, or modification to an existing test because of a new test added to the baseline.

[16] The pie chart only represents the breakdown of number of FT schedule changes and does not take into account the significance or implication of the change.

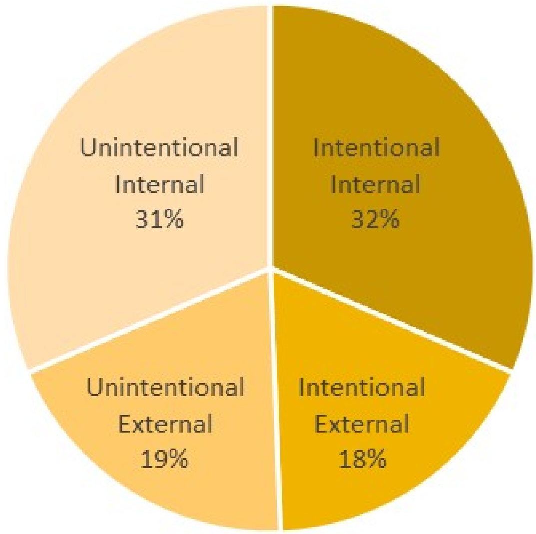


Figure 3-15. Distribution of Flight Test Change Rationales

Figure 3-16 shows FT delays, deletions, or modifications grouped into unintentional and intentional, internal and external changes, with short descriptions listed. The number of tests is measured along the right vertical axis by the orange bar. The blue bars show the average number of days before the test when the change was made, plotted against the left vertical axis.

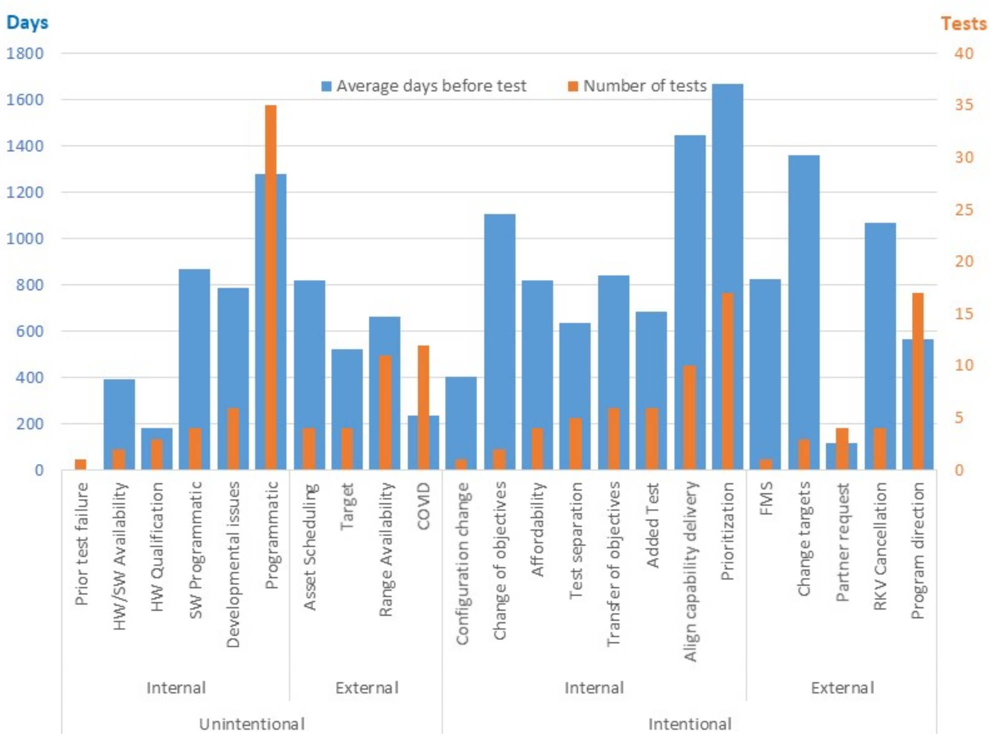


Figure 3-16. Flight Test Change Rationales by Duration and Number

Using this taxonomy, there is no dominant cause for FT schedule changes. However, for unintentional, internal changes that are unforeseen or unplanned but by definition would be within MDA's control to address, Figure 3-16 shows that the largest number of adverse schedule changes are attributed to programmatic reasons, with 35 schedule changes an average of 1,280 days before the FT date.

Because detailed FT planning begins approximately two years ahead of the FT date, blue bars in Figure 3-16 greater than 730 days may not be as consequential as changes that happen within this two-year interval when resources are presumably being expended to plan and execute the test. A particularly undesirable situation, for example, is COVID, which affected a number of tests and the average time before the test date was within a year. COVID was the largest reason for unintentional, external FT changes.

Because COVID affected the last IMTP examined, V22.1, the team also examined the trends of FT change rationales over time. Intentional changes were roughly constant over the past three IMTPs, while COVID shifted the balance between intentional and unintentional changes (53 to 49 percent). If COVID is ignored, 73 percent of unintentional changes are internal and 69 percent of those are programmatic. Figure 3-17 shows the FTs rationales by year using the taxonomy. Note that rationales from V18.0 and V18.1 are combined into a single bar for the fiscal year.

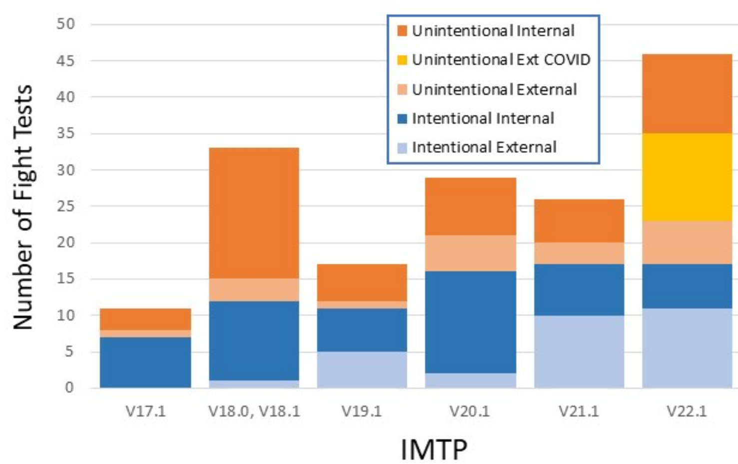


Figure 3-17. IMTP V17.1 to V22.1 Flight Test Delay Reasons by Year

There is no single dominant cause for FT schedule changes, but the programmatic reasons make up 50 percent of unintentional changes excluding COVID. MDA could potentially elaborate on the programmatic rationale,[17] with the aim of using the additional detail to address potential underlying causes. From the timing, programmatic schedule changes appear to be made before detailed planning begins and resources start to be expended on the test, but understanding what they are may help the agency determine if there is a way to reduce them and build a more stable FT schedule.

F. Flight Test Goals

The IDA team examined how FT schedule changes affect the accomplishment of FT goals. A discussion of the purpose of flight testing is detailed in a different chapter; for quantitative analysis, the team looked at how delays in FTs might affect demonstration of BMDS capability increments and how data needs associated with each FT are being measured.

From IMTP V14.2 to present, the BMDS increment being supported by each FT is identified, although the content of each increment can change over time and sub-increments could have been added.[18] For example, Increment 6 was divided into 6A, 6B, and 6C; and then 6B into 6B.1 and 6B.2; and 6C combined with 7. IDA looked at both the average date of the FTs in each IMTP, as well as the date of the last FT supporting each increment. The average date is a proxy measure of when the increment has been partially tested, while the maximum date is a proxy for when the testing for that increment might be complete. The results are plotted in Figure 3-18, and both graphs show schedule delays.

[17] MDA's characterizations of "programmatic" was insufficient for IDA to extract additional detail of underlying causes.

[18] Sub-increments can be added as a result of development or test delays (so some capabilities could be deployed without waiting for others that were experiencing delays), so such break points themselves can be correlated with or used as an indicator of delays.

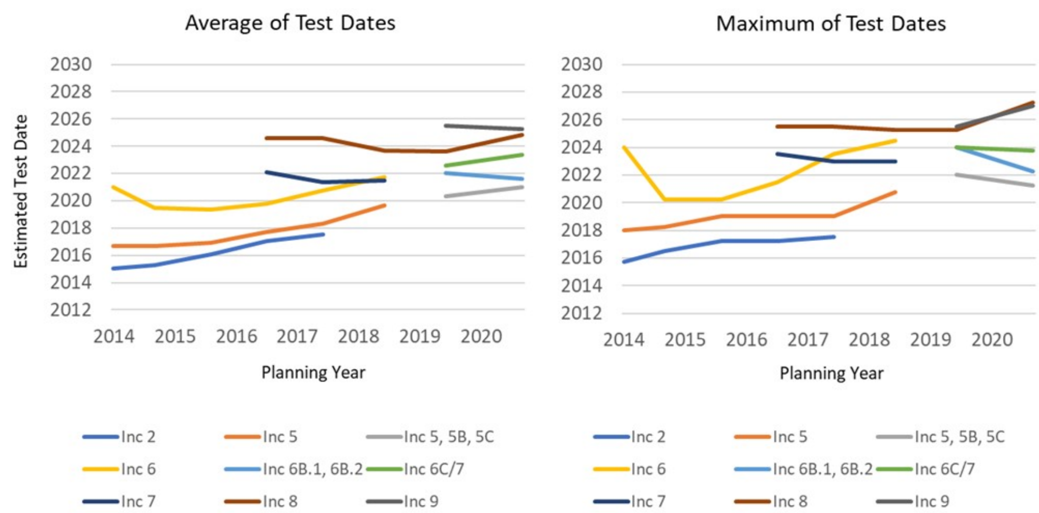


Figure 3-18. Test Dates by Increment

Table 3-4 shows, for each BMDS increment, the number of IMTPs that support that increment and the average test date schedule change per IMTP revision. The maximum total delay is 1077 days (5 revisions x 215 days), or 36 months, for Increment 5, while some increments show acceleration such as Increment 7, moving a total of 7.5 months earlier. Two-thirds of the capability increments show delays in the average FT date. Thus, the IMTPs show that flight testing to demonstrate capability increments is experiencing schedule delays.

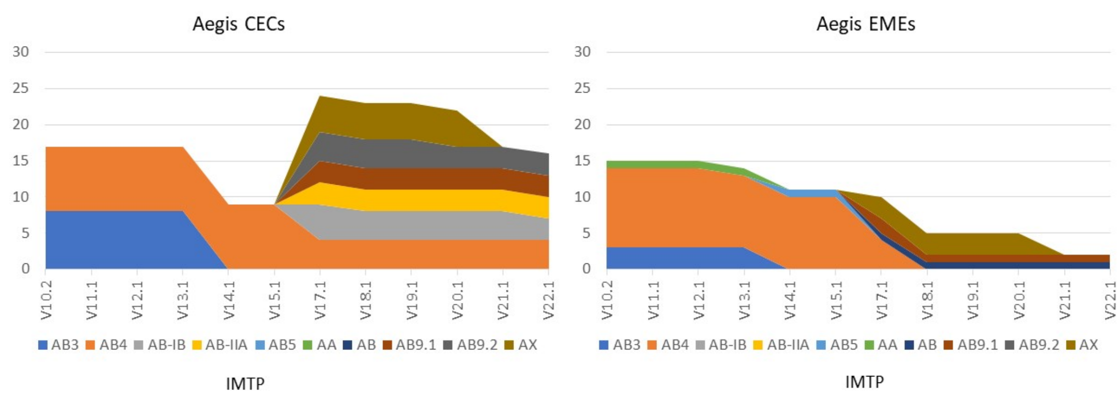
Table 3-4. Flight Test Changes by Increment

	Number of IMTPs supporting increment (#)	Change in Avg Test Date per IMTP revision (days)
Increment 2	5	223
Increment 5	6	215
Increment 5, 5B, 5C	2	239
Increment 6	6	51
Increment 6B.1, 6B.2	2	-156
Increment 6C/7	2	282
Increment 7	3	-113
Increment 8	5	22
Increment 9	2	-92

From Figure 3-18, Increment 6 also shows FT date acceleration in 2014 before the dates begin to be delayed. Each increment adds new capabilities and it is possible that the schedule delays begin as challenges are identified and the planning becomes more informed or because it takes longer than anticipated for the new capabilities to be ready for testing.

To address the collection of data from FTs, the study team examined the reporting of CECs and EMEs in the IMTPs.[19] MDA had previously used CECs and EMEs to manage data collection from FTs, but these metrics have not been listed by FT or in detail since IMTP V20.1.[20] Even when they were listed along with supporting FTs, there was an inconsistency between the IMTP and the overall count in the Smart Books. The Smart Book summaries as of V22.1 still have approved CECs and EMEs by number and element. One IMTP states that all C2BMC EMEs/CECs have been closed or withdrawn and to check the Test Resources Mission Planning Tool (TRMP-T) database for the closure history of each. IDA did not have access to TRMP-T, so the study team used the Smart Book summaries as a source for evaluating overall numbers.

Figure 3-19 is a graph showing the number of CECs and EMEs for different Aegis increments. The graph shows the closure of CECs and EMEs for Aegis Baseline (AB) 3 and 4. It also shows the evolution of Aegis as baseline capabilities expanded and more CEC and EME requirements were added. Based on the current V22.1, all but two Aegis EMEs are closed, and approximately 16 Aegis CECs remain.



AA – Aegis Ashore, AB – Aegis Baseline, AX – Sea Based Terminal

Figure 3-19. Aegis CECs and EMEs by Increment

---

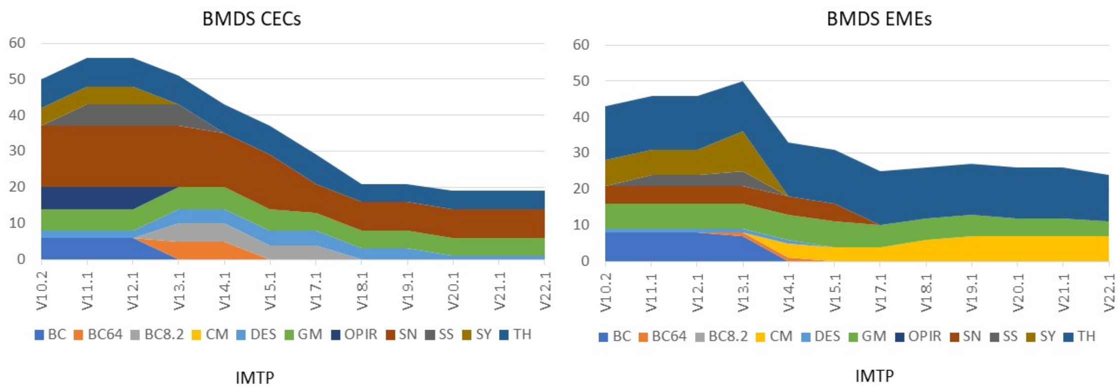
[19] From MDA provided definitions:

Critical Engagement Conditions (CEC): A set of test points that provide data to resolve known M&S uncertainty that limit performance prediction accuracy. The test data collected is used to improve existing models of System or element behaviors. CECs impose conditions on test events that focus data collection efforts with the goal of improving existing M&S

Empirical Measurement Events (EME): A set of test points that provide data for a condition that is not modeled or not modeled to sufficient fidelity. EMEs focus testing on conditions known to produce behaviors that require additional understanding of the operational system, to improve existing low fidelity models or develop new models. As M&S fidelity increases, EMEs may be elevated to CECs

[20] MDA's use of CECs and EMEs is discussed in more detail in Chapter 4.

Figure 3-20 is a graph showing the CECs and EMEs for the BMDS elements excluding Aegis. The Battle Management (BC, BC64, BC8.2) CECs and EMEs have been closed or withdrawn since V17.1. Secondary (SS) and System Level (SY) data needs have also been closed or withdrawn since V13.1. There are a number of THAAD (TH) and GMD (GM) CECs and EMEs that have endured since V10.2. As of V22.1, seven countermeasure (CM), four GMD, and 13 THAAD EMEs remain open, while one M&S (DES), eight Sensor (SN), five GMD and five THAAD CECs remain open.



BC – C2BMC, CM – Countermeasures, DES – Modeling and Simulation, GM – Ground based Midcourse, OPIR – Overhead Persistent Infra-red, SN – Sensors, SS – Secondary, SY – System Level, TH - THAAD

Figure 3-20. BMDS CECs and EMEs excluding Aegis

There are reasons why some data requirements have been closed while others persist. Some programs rely only on ground and element testing to fulfill their data needs. Some FT support to CECs and EMEs has also evolved over time. For example, the current test configuration sheet for FTT-21 shows no assigned CECs or EMEs and there are none in the IMTPs since V20.1, but this test did have one CEC and four EMEs assigned to it in IMTP V14.1, before the TOs changed. Some TOMs mention supported CECs/EMEs exist at an enhanced security level.

It appears that MDA has been able to draw down some of its data requirements through FTs. The Aegis graphic shows new data requirements arise as capabilities are demonstrated and data needs are fulfilled. The THAAD and GMD programs appear to be maintaining a relatively constant number of data requirements.

The CECs and EMEs provide a means to measure the effect of FT schedule delays. The study team did not have access to the TRMP-T database, which is understood to be in the process of being upgraded. But this, or a similar system, can be used by MDA to understand the effect of FT schedule delays on data requirements.

G. Chapter 3 Findings

This chapter examined MDA's FT schedule revisions and instabilities. The study team found the following.

Finding 1: IMTP V22.1 documents a significant increase in the number of FTs over the annual average, but holds roughly constant the number of flight tests that support specific BMDS increments. There has been an increase in the average number of planned FTs per year across IMTPs since V10.2.

Finding 2: IDA can replicate GAO's estimate of 37 percent for MDA's FT execution rate if all FTs over all IMTPs are counted. The GAO metric considers both execution and schedule churn. IDA also found that, within two years of the original best estimated test date, it appears MDA is executing 54 percent of its planned tests.

Finding 3: About 44 percent of FTs have changes to their completion date over the years they appear in budgets and test plans, with a median time of delay of 548 days from the original to the new test date. For each IMTP, excluding V22.1, the average number of tests with schedule changes and the moving average were fairly consistent at around 33 percent per year.

Finding 4: The number of tests per year, both planned and conducted, and the test-related spending appear to be growing by approximately the same percentage. But by some measures, tests are getting more complex.

Finding 5: Using an IDA-developed taxonomy on detailed records since 2017, about half of FT delays, modifications, and deletions are characterized as intentional and the other half are unintentional. About a third of all delays, modifications, and deletions are due to factors external to MDA. COVID-19 was the largest reason for unintentional external FT changes. There is no dominant reason for intentional test schedule changes, but programmatic reasons are cited for the largest fraction of unintentional internal FT changes, and this category should be further elaborated.

Finding 6: By BMDS increment, the average FT dates are being delayed. CECs/EMEs are not tracked in detail after IMTP V20.1. IDA observed that some BMDS element knowledge needs have closed or been withdrawn, while others remain unfulfilled.

H. Chapter 3 Recommendations

The study team recommends that MDA develop its own taxonomy for describing and monitoring FT changes, and use the results to improve flight test planning and execution. In addition, the programmatic reason should be developed in more detail as this dominates the unintentional internal categorization (that is, changes that are not expected and may be within MDA's purview to address). The study team also recommends measuring progress in reaching FT goals. The CECs and EMEs are not explicitly mapped to tests after IMTP

V20.1. MDA should continue to use these or develop and apply other similar measures to manage FT goal accomplishment.

Finally, there are some additional areas that might be explored:

- The data used by IDA had FT dates by quarter; this analysis could not distinguish between days, weeks, or even month-long delays or changes (e.g., due to weather). If short-term execution year schedule changes are of interest, more detailed data would be needed.
- Another analysis might explore differences between FTs. Some FTs may push technological boundaries (for example, intercept tests) more than others (such as sensor only tests or exercises), requiring more resources, planning and coordination. It may be worth exploring if there are differences in schedule change frequency and rationales based on test type.

(This page is intentionally blank.)

4. Discussion of Causes and Consequences of Flight Test Schedule Changes

A. Overview

The BMDS FT program serves several goals: (1) To demonstrate capability and functionality in the successful execution of a mission, (2) to deter adversaries and inspire confidence in U.S. missile defense, (3) to test and evaluate functionality in a spectrum of required engagements in an operationally realistic manner to determine the capabilities and limitations of the system (operational test and evaluation), (4) to identify potential flaws or shortcomings in the system to guide development of the BMDS and resolve those issues (developmental test and evaluation), and (5) to anchor M&S used to evaluate engagements not flight tested. The approaches to achieve these goals often conflict with one another, and all must be served by the same highly constrained FT program. This chapter discusses the interplay of causes and consequences of FT cancellations, delays, and changes as well as other effects found in the review of MDA's FT processes.

Every FT can serve all of these goals, but with different relative efficiencies. The most mission-success-oriented FT will nevertheless discover any deterministic failures in the basic function and integration of the elements in the test. Every repeated test increases the likelihood of discovering stochastic failure modes. Many tests would be required to explore the full functional parameter space of a system, and, for a variety of reasons, large portions of that parameter space are not even accessible to flight testing. The full gamut of feasible tests would serve to most effectively demonstrate a capability as well as discover any latent issues. MDA's challenge is to serve the multiple goals with a severely limited number, cost, and scope of tests.

Many FTs explicitly have both developmental and operational TOs. Developmental test and evaluation aims to discover potential issues, while operational test and evaluation aims to confirm performance under realistic operational conditions.[21] The small number of FTs creates a competition between these two sets of objectives because there are not enough tests to adequately satisfy both. Developmental test and evaluation benefits from exploring the edge of performance space to challenge the system under test. Operational test and evaluation benefits most from a focus on the most likely operational scenarios.

[21] Operational testing can also discover problems not revealed by developmental testing because the latter is often conducted under unrealistic conditions or because the system under test is not fully representative of the operational configuration.

This creates an inherent tension between multiple goals of the FT program. Additionally, deterrence goals push for FTs most likely to result in successful intercepts, at direct odds with attempts to discover unknown issues.

We note that some test programs in industry have taken aggressive approaches with heavy early emphasis on exposing flaws and issues in designs. Although potentially effective, MDA may not have latitude to shift in this direction for legacy programs relied upon for deterrence. However, programs for new or replacement systems may be able to take advantage of some aspects of such approaches before they mature into established deterrents.

B. Demonstrating BMDS Capability

A principal purpose of the BMDS FT program is to demonstrate BMDS capability and verify functionality to support PIP increments and alliance obligations.[22] MDA produces an Integrated Master Assessment Plan (IMAP) documenting all of the assessment goals necessary for the BMDS. FTs represent a unique opportunity to confirm successful implementation of planned capabilities or to discover implementation issues and unmodeled performance issues. However, not all assessment goals can be accomplished via flight testing and many assessment questions are assigned to other means of resolution such as ground testing, component or element testing, or analysis. According to one interviewee, fewer than 5 percent of requirements are allotted to FT verification. A number of interviewees stated that assessment needs do not drive FT planning. Rather, assessment goals are assigned to existing FTs opportunistically where FT collection goals can be adjusted to accommodate the assessment needs or would already provide the necessary data. This suggests that MDA does not intend exhaustive verification of capabilities via FTs for fielding decisions nor could MDA do so, even if desired. Instead, basic functionality is demonstrated in FTs with exploration of engagement parameter space done in ground testing M&S or not at all. However, as will be discussed in greater detail, flight testing serves to anchor the M&S used to address many of the non-flight-tested assessment goals.

Given the need to demonstrate BMDS capabilities to support the PIP, cancellations and delays in flight-test-based system verification force a choice to delay the PIP or devolve to lower-confidence verification methods if requirements cannot be pushed onto other tests and resources. A third alternative would be to release increments of the PIP without full verification, however, according to the interviews, this rarely, if ever, occurs. Delay of the PIP due to FT delays or cancellations occurs only infrequently and is highly undesirable because the PIP represents MDA's public commitment. The most common

[22] The 2019 Missile Defense Review describes the DoD's responsibility and obligation to protect its allies' territory and forces in light of the increasing proliferation of ballistic missiles.

responses to FT delay or cancellation are to move verification requirements onto other FTs where possible and, often, to downgrade verification to lower standards such as ground testing, M&S or analysis. Such downgrading is not accompanied by justifications for a number of reasons. First, there is no standard and no attempt at quantification of verification confidence. Second, in many cases, verification by FT in the first place was opportunistic rather than by design or requirement. Third, there is no formal tracking of confidence down-grading. As a result, consequences of FT cancellations or delays on system demonstration are not prominently observable.

Thus, FT delays and cancellations will likely result in lower-confidence PIP verification. Interviews suggest that downgrading verification needs away from FTs is frequent. Our analysis supports this finding by the fact that, despite FT delays and cancellations, FT data collection needs do not seem to increase with time. When FTs are necessary for particular verification points, switching to other means can be problematic. For verification points served by other means, where flight testing represents an optional confidence boost to be exploited as opportunity allows, thwarted opportunities due to cancellations or delays are unfortunate but not inherently problematic. Nevertheless, the resulting verification is lower confidence than it would have been with flight testing.

MDA has challenges associated with its current process for tracking assessment and verification requirements and progress toward verification goals. This process needs a feedback loop to confirm which verification goals were satisfied and by which means. Issues discovered in a test should be traceable forward to assessments drawn from that test. In the current process, the feedback loop for requirements is informal and ad hoc. There is extensive forum-based collaboration but no single source of authoritative data. Those developing assessment goals meet with engineers to turn those goals into TRs. Those building TRs meet with test designers. Test assessment teams meet with stakeholders and present test results and their implications to MDA leadership. But these meetings mostly do not generate traceable products. Although databases such as TRMP-T exist, passing of knowledge in some of these steps is done via memoranda or presentations, some of which exist in many versions or are not distributed to all relevant parties. The IMAP lists "must haves," while detailed assessment plans have the "like-to-haves." TOMs document the intended goals of each test and often go through multiple iterations. The lack of a single repository accessible to all of the relevant stakeholders and participants makes it difficult to keep track of mapping of requirements to tests and back. Realized, unrealized, and modified verification goals are especially difficult to track in the current process. Another consequence of the lack of a repository is the difficulty in producing a rigorous top-level view for leadership of progress and projected future progress toward meeting assessment and system verification goals.

Sometimes, changes in the test plan occur in response to changes in the PIP. When MDA's declared capability roll-out changes, one would expect the plan for demonstrating those capabilities would change accordingly. Updates to the PIP occur on an as-needed basis in response to real-world events or changes in military priorities. Therefore, such changes can arise anytime within the IMTP development cycle. When the PIP changes with an IMTP update in progress, it can be challenging to ensure that the resulting IMTP adequately meets the verification needs of the updated PIP. Several interviewees noted a need to better align the PIP and the IMTP to help ensure that the current IMTP reflects the needs of the current PIP. The MDA stakeholder community should recognize that a formal document like the IMTP that is released on a regular budget-aligned cycle will necessarily experience some time-lag with respect to changing drivers. It would not be reasonable to expect an IMTP to perfectly reflect all real-world changes occurring after the start of its development cycle. Nor would it be reasonable to expect that real-world changes would occur only in concert with the IMTP development schedule.

C. Characterizing BMDS Capabilities and Limitations

Capabilities and limitations are complementary characterizations of a system. If one thinks of a multi-dimensional operational parameter space (engagement geometry, threat type, closing velocity, solar angle, etc.), MDA would like to identify in which regions of that parameter space a system will perform as required. Capabilities represent the region where the system will perform as required and the associated description of performance. Limitations refer to the regions of parameter space in which the system will not perform as required and the associated description of the manners in which the performance will fall short of requirements. The goal of operational test and evaluation is to identify whether the region of capability covers the required region of capability when the system is used realistically, and where it falls short.

MDA is limited in the number of FTs that can be performed. Success-oriented testing can skirt or delay the discovery of limitations by narrowly exploring the region of parameter space most likely to succeed and then by increasing the parametric region of exploration in small increments. Note that some regions of parameter space cannot be explored due to range or other testing-specific constraints. MDA attempts to explore these regions and other excursions from tested cases as well as possible using ground testing and M&S.

In addition to inherent limitations on the parametric region of capability, there may also be flaws in design. The test and evaluation program attempts to identify these flaws so that they can be fixed or at least understood in terms of their consequent operational limitations. No design is perfect, and the BMDS will likely have design deficiencies. No test program can guarantee discovery of all flaws, and MDA FT program and test and evaluation program as a whole are no exceptions. The purpose of developmental test and evaluation is to efficiently discover these flaws, subject to constrained test opportunities and resources, so that they can be remedied. Some flaws are deterministic and will inevitably disrupt proper function under the right circumstances. Others are stochastic and, even given the right circumstances, will only manifest with some probability. Tests do not prove the absence of flaws, but every test adds confidence in the proper function of the system when flaws do not manifest. The greater the number of tests, the lower the probability that a stochastic flaw will go undiscovered.

FTs are an essential part of this process. FTs can and do find design flaws that may have escaped detection in ground tests and M&S. Complex system flaws can arise in the integration of many subsystems. Unknown unknowns cannot be incorporated into the models used for testing in simulated environments. FTs are perhaps the only way to discover unknown unknowns associated with the integrated system operating in its intended environment, and they can expose unmodeled behaviors of subcomponents in the launch and space environments. Even the most limited FT program will discover the most damaging and basic type of flaw – that which would occur in every flight and would prevent proper function in every use of the system.

Issues discovered during FTs can result in lengthy delays that can disrupt overall FT plans. Examples referenced by one interviewee included FTG-6, FTG-6a, and FTG-7.[23] It is thus far preferable to discover issues prior to the FT via means such as component or ground testing or in M&S. Integration issues can sometimes be discovered during pre-test preparation and integration. When discovered, such issues are generally added to the system and mission assessment plan. Issues discovered in pre-test preparation can delay a test if the time margin is inadequate and can lead to test cancellation if the issues cannot be resolved within the test window. Appendix B discusses optimization of pre-flight schedule margin. Note that some issues are not with the system under test. Sometimes delaying issues are discovered with sensors, range equipment, or targets. These do not reflect on the capability of the BMDS but do impede exploration of the BMDS capabilities and limitations.

D. Deterrence and Warfighter Confidence

An important function of the BMDS is to deter adversaries from using ballistic missiles. Adversary confidence in the successful function of the BMDS is essential for this function. It is also important for U.S. and allied warfighters to have confidence in the BMDS. Successful FT missions bolster both of these functions, and unsuccessful FT missions undermine them. The effect of cancellations and delays of FTs depends on the cause. Delays or cancellations due to delayed development or freshly discovered issues in the system under test make the system seem less robust and undermine confidence,

[23] Two successive CE-II failures (FTG-06 and FTG-06a) resulted in a 2-year FT hiatus before a non-intercept test of another CE-II. The following CE-I intercept test (FTG-07) also failed.

although far less than failures would. Thus, from a deterrence and warfighter confidence perspective, it may be better to incur a delay than to execute a FT at elevated risk of failure. Delays due to shifting MDA or range priorities, weather, target availability, or funding do not inherently undermine confidence in the BMDS, but they do postpone or reshuffle the confidence-building effect of the tests.

Although successful FT missions are essential for building confidence in the system, FTs that seem too easy will detract from the level of inspired confidence. On the other hand, little could diminish confidence in the system more than failed FT missions. One could argue that failures at complex or challenging FT parameters would hurt confidence less than failures at simpler or less challenging tests. Even setting aside other FT goals, which may favor more challenging tests, there is a balance to be struck between the level of challenge posed by FTs and the risk of FT mission failure. Aside from complexity or challenge, there is also a dimension of operational realism. Operational realism may be aligned with complexity or challenge in many cases and certainly aligns with positive contribution to confidence. The team noted MDA's current approach favors scenarios with low FT risk rather than high operational probability or challenge.

One dimension of operational realism posing a challenge to adequate flight testing is threat evolution. Development of FT targets is a lengthy process. Target preparation time can be a major driver of test scheduling. The intelligence community estimates of ballistic missile threats evolve in a shorter cycle than target development. Therefore, the selection of targets in FTs can often unavoidably lag current threat projections. Although this can undermine test realism and detract from system confidence, there is no ready remedy. MDA has made strides in target modularity that can shorten the timeline for incorporation of some threat evolution. MDA attempts to compensates for this threat/target mismatch in M&S, whether by modeling more current threats[24] or via threat design excursions. Unfortunately, such tests are less visible and less reliably accurate than FTs.

E. Support of M&S VV&A

Historically, M&S was a driver of MDA's FT program. Due to the inability to exhaustively test the function of the BMDS in FT, MDA established a strategy to efficiently use the limited FT program for verification, validation, and accreditation (VV&A) of the M&S. This would enable virtual testing of the complete scope of the intended use of the BMDS. The VV&A needs that were to be met using the FT program were divided into two types of knowledge, CECs and EMEs. CECs represent uncertain regions of M&S parameter space needing experimentation to provide the necessary tuning and validation. EMEs represent regions of parameter space where the M&S is not expected to give reliable

[24] Threat model development can be slower than adversary threat development due to a variety of reasons including lack of intelligence on the threat systems.

results and where empirical measurements are required to extrapolate from regions of M&S validity.

CECs and EMEs were once high-profile metrics and drove much of MDA's FT program. However, many interviewees reported that over time these metrics have waned as a driver. Formerly, CECs and EMEs were cataloged, assigned to test events, and tracked system-wide, including their closure. In this fashion, every FT had a concrete place in the VV&A plan, and progress toward full VV&A of the M&S could be measured. The timing of projected validation could readily be planned and predicted, and compared with M&S needs with regard to capability verification for the PIP. Many CECs and EMEs have been closed, but careful tracking is no longer a focus of planning or assessment of FTs.

FTs are still used to update, validate, and accredit M&S for the BMDS. However, less emphasis is now placed on the use of FT data for M&S VV&A. Previously, post-FT reconstruction was used extensively to confirm that model outputs represent observed performance. The interviews suggest that the use of post-FT reconstruction has significantly diminished. Most current use of flight testing for M&S VV&A is opportunistic using available test data. In contrast, in prior practice the M&S VV&A needs were integral to test selection and design with extensive preplanning of data acquisition.

In current practice, ground testing often uses M&S not accredited by FT data. Ground tests may be anchored with FT data that may not be fully threat-representative with threat model excursions conducted to better capture real threats – to some degree this is unavoidable due to the time required to develop or modify test targets following evolution of the real-world threat or assessments thereof. When capabilities and limitations of the BMDS are presented to leadership, use of unaccredited M&S must be declared but is not prohibited. Unaccredited M&S is used routinely, although typically not emphasized, and the associated declaration of capabilities and limitations is routinely accepted by leadership without explicit caveat.

Delays and cancellations of FTs as well as de-scoped tests further diminish or delay the validation of M&S, encouraging the use of M&S without full VV&A[25] and exacerbating the dependence of system acceptance decisions on unaccredited modeling. As a result, declared capabilities and limitations are subject to greater uncertainty and capabilities are deployed with lower confidence. Some interviewees reported that warfighters would rather deploy missile defense systems with lower confidence capabilities or limited scope of confidence in capabilities than wait for more thoroughly vetted capabilities. This problem may be particular to MDA where BMDS evolution lags the threat presenting warfighters with a choice between unattractive options.

---

[25] MDA also explores the use of alternative and creative methods of achieving validation without flight test data. For example, using satellite tracks instead of missile tracks to validate certain radar functions.

(This page is intentionally blank.)

5. Conclusions and Recommendations

A. Conclusions

IDA found the FT process MDA actual executes largely follows the documentation, with stakeholders generally speaking positively of the process. In addition, the team found that the heavily iterative nature of generating an IMTP currently requires repeated instances of substantial input and coordination from numerous stakeholders. Delayed coordination can lead to disruption of the test planning process, which was one of the more common frustrations voiced by interviewees. IDA also observed that, although operational testers external to MDA are engaged early in the overall test planning process (i.e., IMTP), they are often involved late in the detailed planning process for a specific test.

IDA was able to replicate GAO's estimate (37 percent) for MDA's FT execution rate when counting all FTs over all IMTPs. When the team considered changes made within two years of the original best estimated test date, it appears MDA is executing 54 percent of its planned tests. IDA notes that the number of FTs per year and the total test-related spending appear to be growing by approximately the same percentage. But by some measures, the systems under test and the scenarios designed to test them are getting more complex.

GAO's report listed several possible root causes for FT delays and cancellations. IDA found there were also several additional causes. Any related evaluation of MDA should account for the causes of delay, the degree to which they could or should have been avoided, and their consequences. Using an IDA-developed taxonomy, the team found that about half of FT delays, modifications, and deletions are characterized as intentional and the other half are unintentional. About a third of all delays, modifications, and deletions are due to factors external to MDA. Although IDA found no dominant reason for test schedule changes, programmatic reasons are cited for the largest fraction of unintentional internal FT changes, and that this category should be further developed to more clearly identify the causes. Cascading delays, where one delayed test postpones successive tests in a chain reaction, might be avoidable with lower test cadences. However, trading lower test cadence for lower probability of cascading delays does not provide an obvious net benefit to MDA, and delays might sometimes be the lesser price to pay. Where cascading schedule disruptions are a possibility, MDA should certainly account for possible failures or delays and attempt to build resilience into the schedule.

The BMDS FT program serves several goals: (1) Demonstrating BMDS Capability, (2) Characterizing BMDS Capabilities and Limitations, (3) Deterrence and Warfighter

Confidence, and (4) Support of M&S VV&A. These goals often conflict with one another, and all must be served by the same highly constrained FT program. IDA examined the interplay of causes and consequences of FT cancellations, delays, and changes as well as other effects found in the review of MDA's FT processes.

When demonstrating BMDS capability and verifying functionality to support PIP increments, cancellations and delays in flight-test-based system verification often force MDA to devolve to lower-confidence verification methods. IDA found that such downgrading is not accompanied by traceable justifications and as a result, consequences of FT cancellations or delays on system demonstration are not prominently observable. MDA also has challenges associated with its current process for tracking assessment and verification requirements and progress toward verification goals. This process has need of a feedback loop to circle back after testing to confirm which verification goals were satisfied and the means by which they were satisfied to track progress and project future progress. Historically, M&S was a driver of MDA's FT program. Due to the inability to exhaustively test the function of the BMDS in FT, MDA established a strategy to efficiently use the limited FT program for VV&A of the M&S. IDA found that less emphasis is now placed on the use of FT data for M&S VV&A. Delays and cancellations of FTs as well as de-scoped tests further diminish or delay the validation of M&S, encouraging the use of M&S without full VV&A and exacerbating the dependence of system acceptance decisions on unaccredited modeling.

B. Recommendations

IDA recommends that MDA consider the following adjustments to its current approach.

- Pre-flight testing is a good opportunity to discover potential issues. MDA should conduct a trade-study to assess whether adjusting FT pre-mission schedule margins would produce a net benefit.
- MDA should review the IMTP generation process and reconsider if stakeholder coordination is truly required in each of the IMTP process steps currently noted (2/3 of the 34 steps) and streamline the process as appropriate.
- MDA should utilize operational testers earlier in the detailed FT planning process, ideally beginning with the Kick-off Planning Meeting.
- MDA should develop its own taxonomy for describing and monitoring FT changes and use the results to improve flight test planning and execution.
- MDA should develop a single-source traceable end-to-end mapping of assessment objectives for tests to evaluate progress toward meeting verification

goals. The process should include the use of CEC/EMEs[26] or similar but more comprehensive measures to manage the accomplishment of FT goals and include a feedback loop that circles back after testing to confirm which verification goals were satisfied and the means by which they were satisfied.

- MDA should establish a formal tracking procedure to better understand confidence down-grading when relying on lower-confidence verification methods (i.e., non-FT based) due to cancellations and delays in flight-test-based system verification.

[26] CEC/EMEs were high-profile metrics and as such were encouraged to be kept to a manageable number. Similar but lower-profile measures could potentially be much more comprehensive.

(This page is intentionally blank.)

Appendix A. Interviewees and Interview Dates

Interviewee		Date
MDA Director of Engineering (DE)		
Dave McNeil	System Engineer	7/26/21
Brian Godwin	Director Systems Engineering Test Requirements (SEF), Phase I OPR for JETID	7/22/21
Steve Peduto	SEF Lead Integrator	
Jeffrey King	Systems Engineering Test Requirements (SEF)	
Kathleen Miga	Director System Design and Specifications (SED), publishes PIP	7/22/21
Jeff Stidham	Director Systems Assessment (SEG), publishes IMAP	7/16/21
MDA Director of Test (DT)		
Kevin Williams	Technical Director	7/27/21
Bob Croft	Director Test Planning and Design (DTD)	7/19/21
Ashley Garrison	Deputy Director Test Planning and Design	7/19/21
Marlon Thompson	IMTP Lead (DTDI)	7/19/21
Chris Brown	IMTP Flight Test Lead	
Ken Ross	Test Baseline WG Secretary	
Kurt Lambert	IMTP Publisher, Historian, and Contract Lead	7/23/21
Tim Smith	Test Baseline Manager	7/28/21
Melissa Long	Director for Test Budget (DTB)	7/30/21
Bruce Bush	DTB Test Lead	
Michelle Nelson	Director for Test Resources (DTR)	8/5/21
Jim Ussery	Director for Flight Test (DTF)	7/28/21
Stephen Beard	Deputy Director for Test	8/10/21
Mike Morrissey	Director for Test (DT)	
Ronald Yuhasz	Director for Test (DT)	
MDA Director of Operations (DO)		
Alexandria Martinez	Comptroller	8/27/21
Michelle Atkinson	Director for Operations (DO)	8/6/21
MDA Targets and Countermeasures		
Dave Moriarty	Deputy Director for Targets and Countermeasures	8/9/21
Shaun Lee	Targets and Countermeasures	
MDA Threat Engineering		
Abe Bushra	Director Threat Systems Engineering	8/4/21
Stakeholders and Others		
Steven Lopes	DTE&A	8/6/21
Phil Clark		
Bill Swank		
Mitch Crosswait	DOT&E	8/18/21
Brian Jones		
Mike Gilmore	IDA - SED	8/10/21
Rick Mraz	IDA (MDIOC Support)	8/9/21
Mark Briski		
Tom Schorsch		
Dave Richie		
Kirk Olson		
Mike Luhman	IDA - OED	8/13/21
Dawn Roper		
Robert Salow		
Swati Varshney		
Quentin MacManus	JFCC-IMD	9/10/21
Ian Muntean		
Ryan Smith		
Michael Christensen	AFOTEC	8/18/21
Anamaria Dent		
Todd Alexander	BMDS OTA	
Mark Jahnig		
Chris Atkinson	COTF	8/26/21
William Phillips	ATEC	9/2/21
Jonathan Nolan		
Gene Nixon	NORTHCOM	9/16/21
Jeff Dillemuth		

(This page is intentionally blank.)

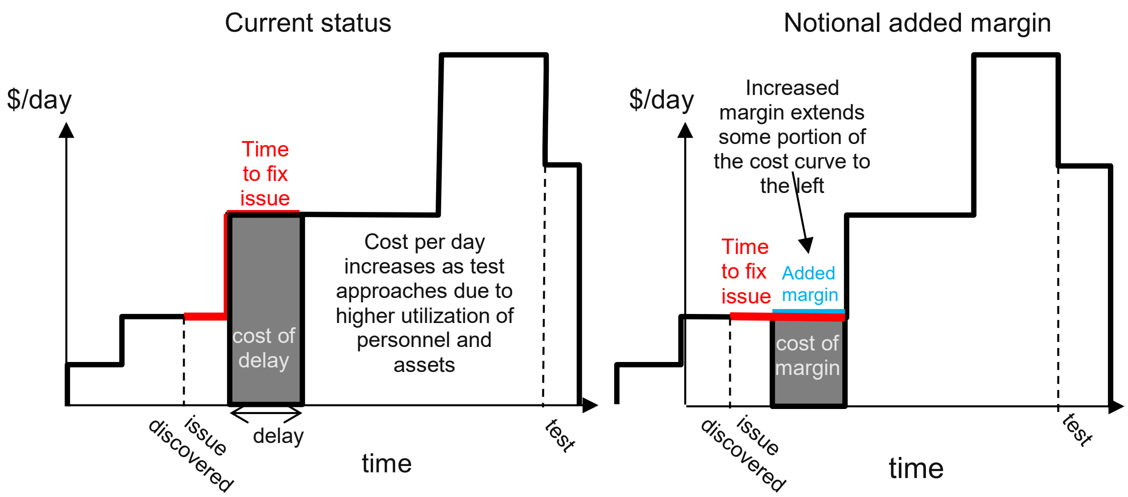
Appendix B. Analyzing the Cost and Benefit of Pre-flight Test Margin

Pre-flight testing is an opportunity to discover potential issues that could arise during flight testing. Sometimes, when issues are discovered during pre-flight testing, their resolution requires either delaying or cancelling a FT or flight-testing at greater risk. Shifting the pre-flight testing period earlier would provide a greater margin of time for fixing issues thus identified. Additional schedule margin would also provide opportunity for extended pre-flight testing. On the other hand, adding schedule margin in this fashion would entail additional cost and challenge, both due to the activities themselves and the extended deployment of personnel. It would need to be determined how, if at all, personnel, range, and equipment could be made available. However, FT delays also increase costs for the same reasons and introduce scheduling and test resource challenges, including possible cancellation of tests. A more predictable schedule would reduce burdens on deployed personnel and operational assets. FT delays also decrease confidence in the test and evaluation enterprise, as evidenced by the GAO report which led to this study.

A priori, the optimal additional pre-FT schedule margin is not clear, however, MDA should have access to historical data on the frequency and timing of delays due to pre-flight issue discovery, as well as the associated costs of both delays and of pre-flight testing. Note, IDA did not analyze whether pre-flight issue discovery represents a significant contributor to flight-test delays, cancellations, or cost increases. The value of this analysis depends on the degree to which such issues occur.

Figure B-1 shows a figurative view of the potential benefit of adding pre-flight schedule margin. The preparation for a test incurs cost. As the test date grows closer, increasing amounts of personnel and equipment are present and active in pre-test activities, and the daily cost increases accordingly. When issues arise, some parts of that escalating cost curve may be postponed while the issues are addressed, and some parts may inevitably escalate. Some portions of the cost curve may be prolonged until the issues are resolved. If the issues can be resolved within the allotted time, the test may proceed as scheduled. If resolution of the issues exceeds the allotted time, the test can be delayed. Delays that push the test out of the available window can lead to test cancellation. Beginning the pre-mission testing and integration process earlier could leave room for additional schedule margin during comparatively low-cost portions of the daily-cost curve. In that case, if issues are discovered early enough, issue resolution may be completed without extending the duration of higher-cost portions of the schedule and without delaying the test. The plots in Figure

B-1 show a notional example of potential cost savings due to this effect. There are important implications other than cost, but cost will be addressed first.



Notional effect of added margin. In the left-hand panel, an issue arises requiring more time to fix than the margin allotted. Consequently, expensive stages of the lead-up to FT become extended resulting in additional costs and a delayed test. The cost of the delay is only incurred when such issues arise. In the right-hand panel, the same issue arises, but due to the availability of additional margin extending the duration of a lower-cost stage of flight-test preparation, the issue is fixed within the new extended time available. The cost of the added margin is incurred whether or not an issue arises requiring use of the margin. Notionally, the margin is added with earlier deployment to the test site without changing the overall test cadence.

Figure B-1. Notational Effect for Pre-flight Test Margin

From a cost perspective, the current level of pre-flight testing and schedule margin may be too high, too low, or just right. The costs of margin and other elements of pre-flight testing are incurred whether or not issues are discovered and additional time is needed. The cost of delays is incurred stochastically depending on the probability of discovery of issues and the associated timing of issue discovery and duration of issue understanding and resolution. MDA should have the data required to do a cost analysis of pre-flight margin – the historical probability and timing of issue discovery, the time required to resolve issues, and the cost per day of the various stages of pre-flight activities. IDA believes it would be worthwhile for MDA to conduct such an analysis, bearing in mind that cost would not be the only deciding factor in the consideration of margin adjustment.

For illustrative purposes, a notional analysis is shown of the type suggested using a highly simplified model. First, consider a cost model for the process with potential delays. Let $P(\tau, t_0)$ and $Q(\tau, t_0)$ be the probability density and average cost, respectively, of a delay of duration τ beginning at time t_0. If the cost with no delays is Q_0 and the duration with no delays is T_0, then the total cost would be

$$C = Q_0 + \int_0^{T_0} \int_0^{\infty} P(\tau, t_0) Q(\tau, t_0) d\tau dt_0 \,, \tag{1}$$

where the final term is the probability-weighted cost of delay.

Currently, MDA's success-oriented approach assumes there will be no delays and the total cost will be Q_0. The team heard a theme in the interviews when raising the question of success-oriented mentality which suggested that to do otherwise would be to plan to fail. Failures can and do occur. Issues can and do arise. Delays can and do occur. MDA might be doing a disservice in assuming success when scheduling and costing the portfolio. Planning for failure is different from planning to fail. Planning for failure is an attempt to introduce resilience, to plan for contingencies. No one has recommended planning to fail. That is a straw-man argument making realism anathema to planning. In planning for failure, MDA could use past experience to judge the likelihood of future success. MDA could plan for success but be prepared for disruptions. That begins with taking a realistic look at the probabilities of delays and their associated costs and other consequences. Even if this information is not used to optimize pre-flight schedule margin, MDA could consider incorporating this kind of probabilistic cost and schedule modeling into its portfolio management.

To investigate the effect of added margin, the team considered a highly simplified model where all delays occurred at a fixed stage of the pre-flight process, and added margin delays at the onset of the added cost function. Let T be the duration of added margin with associated cost M(T). The expected cost with margin incorporated is

$$C = Q_0 + M(T) + \int_0^{\infty} P(\tau + T) Q(\tau) d\tau \,. \tag{2}$$

Looking at the net expected change in cost of adding margin T to the existing approach (the difference between Eqs. (1) and (2)),

$$\Delta C(T) = M(T) + \int_0^{\infty} P(\tau + T) Q(\tau) d\tau - \int_0^{\infty} P(\tau) Q(\tau) d\tau \,. \tag{3}$$

In other words, the cost differential is the difference between the fixed additional cost of the added margin plus the expected (probability-weighted) cost of delays exceeding the margin and the expected cost of delays absent the added margin. The cost of cancellation can be incorporated into this model if delays greater than τ_C lead to cancellation with fixed cost Q_C,

$$Q(\tau > \tau_C) = Q_C \,. \tag{4}$$

The optimal amount of added margin is given by minimizing $\Delta C(T)$.

If, for example, the cost of delay were proportional to the duration of delay, $Q(\tau) = q\tau$, and the cost of margin were proportional to the duration of margin, $M(T) = mT$, then the optimal margin would be given by

$$\int_0^T P(\tau)d\tau = \frac{q - m}{q} \ . \tag{5}$$

The left-hand side of Eq. (5) is the probability of a delay that would impact cost only absent the additional margin. The right-hand side is the relative cost discount associated with time devoted to margin versus the equivalent delay. Interpreting Eq. (5) in this simple model, if the cost per day of margin were X% cheaper than the cost per day of delays, then the optimal margin would be that which accommodated X% of delays. If some margin is already built in, it is quite possible that the optimal amount of margin from a cost-only perspective would actually be lower than at present.

Adding schedule margin during the pre-flight period and/or adding additional pre-flight testing may not require reducing the flight-test cadence, which would be an unpalatable impact for many stakeholders because FTs are already in far lower supply than demand. It could increase the probability of successful FTs, a principal goal of the MDA FT program. It should reduce the likelihood of FT delays and increase stability and predictability of the schedule, a desirable goal for MDA and external stakeholders. There is an open question as to whether it would result in net cost reductions, but MDA could perform the cost analysis to enable carefully considered cost-benefit decisions incorporating a realistic perspective on issues and delays

.

Appendix C. Acronyms and Abbreviations

AA	Aegis Ashore
AB	Aegis Baseline
AEDC	Arnold Engineering Development Center
AFOTEC	Air Force Operational Test and Evaluation Center
ATEC	Army Test and Evaluation Command
AX	Sea Based Terminal
BMDS	Ballistic Missile Defense System
BETD	Best Estimated Test Date
BOA	Broad Ocean Area
BVT	Booster Vehicle Test
C2BMC	Command and Control, Battle Management, Communications
CEC	Critical Engagement Conditions
CM	Countermeasures
COMOPTEVFOR	Commander, Operational Test and Evaluation Force
CTV	Controlled Test Vehicle
DE	MDA Directorate of Engineering
DO	MDA Directorate for Operations
DoD	Department of Defense
DOT&E	Director, Operational Test and Evaluation
DSUT	Developmental Systems Under Test
DT	MDA Directorate of Testing
EME	Empirical Measurement Event
FS/PD	Formidable Shield/ Pacific Dragon
FT	Flight Test
FTCN	Flight Test Communications Network
FTG	Flight Test GMD
FTM	Flight Test Aegis
FTO	Flight Test Operational
FTP	Flight Test Patriot
FTT	Flight Test THAAD
FTX	Flight Test Other

GAO	Government Accountability Office
GM	Ground based Midcourse
GMD	Ground-based Midcourse Defense
HALO	High Altitude Observatory
ICBM	Intercontinental Ballistic Missile
IDA	Institute for Defense Analyses
IETT	Integrated Event Test Team
IMAP	Integrated Master Assessment Plan
IMTP	Integrated Master Test Plan
IRBM	Intermediate Range Ballistic Missile
JFCC-IMD	Joint Functional Component Command for Integrated Missile Defense
KTF	Kauai Test Facility
LRDR	Long Range Discrimination Radar
M&S	Modeling and Simulations
MATSS	Mobile At Sea Sensor System
MD	Mission Director
MDA	Missile Defense Agency
MDA/DT	Missile Defense Agency/Directorate for Test
MDAR	Missile Defense Accountability Report
MDIOC	Missile Defense Integration and Operations Center
MIDCON	Missile Defense Communication and Operations Node
MRBM	Medium Range Ballistic Missile
NSITE	Network System Integration and Test Environment
OPIR	Overhead Persistent Infra-red
OSD	Office of the Secretary of Defense
OTA	Operational Test Agent
OTRR	Operational Test and Readiness Review
PB	President's Budget
PD	Program Director
PE	Program Executive
PIP	Phase Implementation Plan
PM	Program Manager
PMRF	Pacific Missile Range Facility
POM	Program Objective Memorandum
RASA	Remote Area Safety Aircraft

RKV	Redesigned Kill Vehicle
RTS	Reagan Test Site
SBIRS	Space-Based Infrared System
SBX	Sea-based X-band radar
SEIC	Systems Engineering Integration Council
SM	Standard Missile
SN	Sensors
SRBM	Short Range Ballistic Missile
SUT	Systems Under Test
SY	System Level
TC	Targets and Countermeasures
TD	Test Design
THAAD	Terminal High Altitude Area Defense
TO	Test Objective
TOM	Test Objective Memoranda
TR	Test Requirement
TRMP-T	Test Resources Mission Planning Tool
TSS-D	Test Support System Data
TSS-SA	Test Support System Situational Awareness
UEWR	Upgraded Early Warning Radar
VAFB	Vandenberg Space Force Base
VV&A	Verification, Validation, and Accreditation
WSMR	White Sands Missile Range

(This page is intentionally blank.)

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**	
02-2022	Paper	**START DATE**	**END DATE**

4. TITLE AND SUBTITLE
Independent Assessment of Missile Defense Agency Processes for Flight Test Planning and Execution

5a. CONTRACT NUMBER	**5b. GRANT NUMBER**	**5c. PROGRAM ELEMENT NUMBER**
HQ0034-19-D-0001		
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
DB-1-4959		

6. AUTHOR(S)
Buontempo, Joseph, T.; Hong, John, S.; Girardi, Matthew, R.; Uy, Robert, V.; Teichman, Jeremy, A.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
730 East Glebe Road
Alexandria, Virginia 22305

8. PERFORMING ORGANIZATION REPORT NUMBER
P-30853
H 2021-000403

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
COL Bob Croft
MDA Test Planning and Design (DTD)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
One of the main findings of the July 2020 Government Accountability Office (GAO) report “Missile Defense: Assessment of Testing Approach Needed as Delays and Changes Persist” is that the Missile Defense Agency (MDA) “did not fully execute its fiscal year 2019 flight testing, continuing a decade-long trend in which MDA has been unable to achieve its fiscal year flight testing as scheduled. Although MDA revised its approach to developing its annual test plan in 2009 to ensure the test plan was executable, over the past decade MDA has only been able to conduct 37 percent of its baseline fiscal year testing as originally planned due to various reasons including developmental delays, range and target availability, or changing test objectives.” The GAO report recommended that “MDA ensure an independent assessment is conducted of its process for developing and executing its annual [Ballistic Missile Defense System] BMDS flight test plan.” The Department of Defense (DoD) concurred with this recommendation, and MDA asked the Institute for Defense Analyses (IDA) to conduct the assessment. The scope of IDA’s study included the BMDS flight test program and its role in characterizing the system’s specific capabilities and limitations, demonstrating the overall BMDS capability to allies and adversaries, deterring adversaries’ use of ballistic missiles, building warfighter confidence, and validating and verifying the BMDS models and simulations. Other aspects of the BMDS test program – such as ground tests, the use of models and simulations, exercises and wargames – were outside the scope of this assessment. IDA’s overarching approach to conducting the study included reviewing relevant documentation and conducting interviews with government officials from MDA and other relevant DoD stakeholder organizations to hear their perspectives and learn from their experiences. The information was analyzed and common trends and issues were identified. This report documents IDA’s findings and recommendations.

15. SUBJECT TERMS
Ballistic Missile Defense System; IMTP; Missile Defense Agency; Flight Tests

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	
Unclassified	Unclassified	Unclassified		

19a. NAME OF RESPONSIBLE PERSON	**19b. PHONE NUMBER**
Joseph Buontempo	703-845-2349