412TW-TIH-17-02

ENVELOPE EXPANSION LESSONS LEARNED

PAUL SOROKOWSKI
Project Engineer

MICHAEL P. GARLAND
Project Engineer

4
1
2
T
W

MARCH 2018

TECHNICAL INFORMATION HANDBOOK

DISTRIBUTION A. Approved for public release; distribution is unlimited.

412TW-PA-18145

412TH TEST WING
EDWARDS AIR FORCE BASE, CALIFORNIA
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

This technical information handbook (412TW-TIH-17-02, *Envelope Expansion Lesson Learned*) was requested by the Director, 412th Test Engineering Group, Edwards AFB, California 93524 in response to inquiries from project office customers developing new USAF aircraft.

Prepared by:

SOROKOWSKI.PAUL.J.1231882207 Digitally signed by SOROKOWSKI.PAUL.J.1231882207 DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=USAF, cn=SOROKOWSKI.PAUL.J.1231882207 Date: 2017.10.18 15:26:28 -07'00'

PAUL SOROKOWSKI
Project Engineer
773d Test Squadron

Michael Garland (affiliate) Digitally signed by Michael Garland (affiliate) Date: 2017.10.19 14:34:36 -05'00'

MICHAEL P. GARLAND
Project Engineer
Jacobs Technology

This report has been reviewed and is approved for publication: 1 March 2018

CACANINDIN.ARTEMIO.G.1159806371 Digitally signed by CACANINDIN.ARTEMIO.G.1159806371 DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=USAF, cn=CACANINDIN.ARTEMIO.G.1159806371 Date: 2017.10.20 11:22:23 -07'00'

ARTEMIO G. CACANINDIN
Director (Acting), Flight Systems Integration
773d Test Squadron

OSBURN.DANIEL.W.1241451239 Digitally signed by OSBURN.DANIEL.W.1241451239 Date: 2018.03.01 14:57:15 -08'00'

DANIEL W. OSBURN
Director
412th Test Engineering Group

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - Through)*
03-01-2018	Technical Information Handbook	N/A

4. TITLE AND SUBTITLE
Envelope Expansion Lessons Learned

5A. CONTRACT NUMBER

5B. GRANT NUMBER

5C. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Sorokowski, Paul, Project Engineer
Garland, Michael P., Contractor Project Engineer

5D. PROJECT NUMBER

5E. TASK NUMBER

5F. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
412th Test Wing
Edwards AFB CA 93524

8. PERFORMING ORGANIZATION REPORT NUMBER
412TW-TIH-17-02

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
412th Test Engineering Group
307 E. Popson Ave.
Edwards AFB CA 93524

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES
SC: 012100 CA: 412th Test Wing Edwards AFB CA Print this document in Black and White

14. ABSTRACT
This handbook documents recurring issues and best practices that have been observed during the air vehicle envelope expansion phase on multiple USAF flight test projects over at least three decades. Recommendations are made on methods that have been used to alleviate the negative impacts of each issue and emphasize the positive aspects of the best practices. These lessons are the result of a collaboration of more than a dozen flight test engineers all highly experienced with envelope expansion. There was no attempt to capture lessons learned from the mission systems (avionics, electronic warfare, etc.), or sustainment portions of flight test projects although several of the lessons would also be applicable.

15. SUBJECT TERMS
Envelope expansion, lessons learned, best practices, Flying Qualities flight testing, Flutter flight testing, Loads flight testing, Air Vehicle flight testing, first flight, simulation, flight test aids, residual oscillations, rate limited control surfaces, decision making, collaboration, coordination, project management

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19A. NAME OF RESPONSIBLE PERSON / 19B. TELEPHONE NUMBER (INCLUDE AREA CODE)
U	U	U	Same as Report	340	773 TS/ENF / 661-277-4138

STANDARD FORM 298 (REV. 8-98)
PRESCRIBED BY ANSI STD. Z39.18

This page was intentionally left blank.

AUTHORS BIOGRAPHY

PAUL SOROKOWSKI

Paul Sorokowski is an aerospace engineer with over 41 years flight test experience, almost all of it associated with envelope expansion on leading edge aircraft designs. His first exposure to envelope expansion concepts was in the 1970s as a performance engineer during Full Scale Development (FSD) for the F-15 and F-16 as well as the YC-15 prototype. That was followed by the Advanced Fighter Technology Integration F-16 project in the early 1980s. The AFTI/F-16 was a joint research project with NASA to evaluate cutting edge technologies that were later applied on several production fighters. He was very much part of the core team developing multi-discipline envelope expansion concepts as the lead flying qualities/flight controls engineer on the B-2 project during most of Engineering and Manufacturing Development (EMD). He reprised that multi-discipline envelope expansion role throughout all of F-22 EMD. A different set of envelope expansion concepts was developed during the Automatic Ground Collision Avoidance System (Auto GCAS) F-16 Fighter Risk Reduction Project, which was another joint project with NASA to mature the Auto GCAS technology. More recently, during the very early stages of a major new USAF EMD project he was given the opportunity to influence contractual language that captured some of the key concepts proposed in this handbook.

MICHAEL GARLAND

Michael Garland is an aerospace engineer with over 41 years flight test experience, much of it associated with envelope expansion on leading edge aircraft designs. His first exposure to envelope expansion concepts was in the 1970s as a performance engineer on the YC-15 Advanced Medium STOL Transport (AMST) and then as a flying qualities engineer during FSD of the F-16. The F-16 work continued into the F-16 production Block upgrade efforts. That was followed by the F-16XL Multi-Role Fighter Program in the early 1980s. Subsequent to the F-16XL, he became the lead flight dynamics engineer on the B-1B development program. This led to becoming the Deputy for Engineering and then the Deputy Director at the F-16 Combined Test Force (CTF). From the F-16 CTF, he became the Deputy Director of the B-2 CTF during the B-2 EMD. He then supported the stand-up of the Electronic Warfare Directorate as the Chief of the Electronic Warfare Test Division and then the Deputy Director of the EW Directorate. In 2005 he became the Director of the 412th Test Engineering Group. He retired from Air Force civil service in 2009 and became a contractor supporting NASA on the Auto GCAS F-16 Fighter Risk Reduction Project, to mature the Auto GCAS technology. He subsequently supported the development of the F-16 Airborne Collision Avoidance System (ACAS), and the integration of ACAS and GCAS in the F-16 Integrated Collision Avoidance System (ICAS) development. Other recent efforts include supporting development of the use of run time assurance technology to enable autonomous aircraft operations in the National Airspace.

This page intentionally left blank.

PREFACE

During the early research to identify which flight test lessons were appropriate to include in this compilation, one of the first steps was to interview a number of flight test engineers that were highly experienced in envelope expansion of new aircraft. Those interviews established the overall scope for this compilation. The authors would like to thank the following individuals for providing their valuable expertise and helping identify the most important topics:

John Manke
Robert Lee
Kathy Wood
Mark Crawford
Tim Cacanindin
Brian Hobbs
Jessica Peterson
Chris Eaton
Fred Webster
Jason Bostjancic
Kirk Harwood
Reagan Woolf
Wendy Hashii

In addition, the authors would like to thank the following individuals for providing excellent feedback comments during the final review process for this handbook:

John Manke,
Mark Crawford
Tim Cacanindin
Kirk Harwood
Reagan Woolf
Wendy Hashii
Joey Duncan
Firdosh Choksey
Mike Roy
Joe Rowe
Wah Wong
Chuck Webb
Jim Harris
Dave Harmon, JT3 LLC

This page was intentionally left blank.

EXECUTIVE SUMMARY

This handbook was motivated by questions from 412th Test Wing (412 TW) members and customers in preparation for new flight test projects. Those questions can be paraphrased as: "What are the lessons that have been learned during past envelope expansion flight test projects at Edwards Air Force Base?" Responding to that question was more problematic than expected. There were many individuals who could relate their own experiences, but there was no single source that represented the combined experience of many flight testers. This handbook represents the first attempt to provide that compilation in a form that will be accessible via the Defense Technical Information Center and the 412 TW Technical Library. It is fully expected that this handbook will be revised and updated as additional lessons learned are assembled from previous and ongoing envelope expansion projects.

There was no attempt to capture lessons learned from the mission systems (avionics, electronic warfare, etc.), or sustainment portions of flight test projects, although several of the lessons learned included here are not discipline specific so also apply to these disciplines. This handbook primarily includes lessons learned during development programs for new production aircraft, but also includes lessons from research projects to evaluate cutting edge technologies, many of which were later incorporated into production aircraft.

When assembling the lessons in this handbook, there was a conscious effort to avoid picking topics that only occurred on a single project. The guiding principle was to identify recurring themes that seemed to happen regardless of the type of aircraft (bomber, fighter, transport, etc.) or the specific organizations involved with designing and developing those aircraft. Therefore the resulting compilation may not be what some readers expect. These lessons do not directly discuss the discipline-specific methodologies typically associated with the conduct of envelope expansion flight testing for Flutter, Loads, Flying Qualities, Propulsion, High Angle of Attack, etc. The topics selected may address recurring issues related to some of those disciplines. However, the main goal was to capture the top level processes and decisions that have contributed to generic problems which consequently inhibited envelope expansion, and to recommend methods that have enhanced the envelope expansion process.

The lessons in this handbook were developed with the concept of "Early Tester Involvement" in mind and were organized into groupings based on the timeframe when inputs from flight testers would be most effective for improving an overall development effort. Those groupings are:

- Topics to Coordinate Prior to Contract Award
- Topics to Coordinate Prior to First Flight

Flight testers may not always be involved prior to contract award. Without the ability to influence the wording of the Statement of Work or other contractual documents, flight testers will be highly constrained on how much positive impact can be achieved. In those situations, any positive impacts will need to be a result of "common interest" negotiations with the program office and prime contractor. There is no grouping for "Topics to Coordinate after First Flight" because that would typically be too late to be effective with respect to envelope expansion.

Each lesson learned in this handbook is presented in two formats. The main text presents a classic lessons learned format with quite a bit of background and discussion. That format is intended for the more technical members of the project. The second format is presented in Appendix A and consists of a few briefing slides that can be used to communicate the gist of each topic to management or other audiences without getting bogged down in as much technical detail.

This page was intentionally left blank.

TABLE OF CONTENTS

Page No.

This page was intentionally left blank.

INTRODUCTION

"Those who fail to learn from history are doomed to repeat it"

Winston Churchill (and others)

Many people are familiar with that famous quote. However, most people may not be familiar with a longer version:

"When the situation was manageable it was neglected, and now that it is thoroughly out of hand we apply too late the remedies which then might have effected a cure…

…Want of foresight, unwillingness to act when action would be simple and effective, lack of clear thinking, confusion of counsel until the emergency comes, until self-preservation strikes its jarring gong– these are the features which constitute the endless repetition of history."

Excerpts from Churchill's speech to House of Commons, 2 May 1935

Churchill clearly had other problems in mind when he was motivated to communicate those thoughts, but his words could easily be applied to many previous large scale development flight test projects. A number of "situations" have occurred with dismaying predictability on those past projects regardless of the acquisition strategy and regardless of the type of aircraft or the specific organizations involved. The overall goal of this handbook is to identify some of those situations and recommend remedies that have been used to avoid those situations or at least minimize the negative impacts. Every lesson learned may not be applicable to every new project, and there may be other methods that will work as well. It is hoped that people assigned to envelope expansion projects will seriously consider the tradeoffs and be willing to act by implementing project-specific solutions that may help avoid the endless repetition of flight test history.

This handbook captures recurring issues and best practices that have been observed during the air vehicle envelope expansion phase on multiple USAF and NASA flight test projects over at least three decades. Recommendations are made on methods that have been used to alleviate the negative impacts of each issue and emphasize the positive aspects of the best practices. These lessons are the result of a collaboration of more than a dozen flight test engineers all highly experienced with envelope expansion. There was no attempt to capture lessons learned from the mission systems (avionics, electronic warfare, etc.), or sustainment portions of flight test projects although several of the lessons learned included here are not discipline specific so also apply to these disciplines. There was no attempt to capture lessons learned from the unique perspective of test pilots or other aircrew, test conductors, or maintainers.

Most of these flight test related envelope expansion lessons were observed during large scale programs (Full Scale Development [FSD], Engineering and Manufacturing Development [EMD], System Design and Development [SDD]). Some of these lessons were observed during joint USAF/NASA flight test projects on research aircraft to evaluate cutting edge technologies that were later incorporated onto production aircraft. However, many of the lessons would still be applicable to smaller scale envelope expansion projects and some aspects of sustainment projects.

There was a conscious effort not to pick on any one test project. Much emphasis was placed on identifying issues that have recurred on multiple test projects regardless of the specific aircraft type, program office, test organization, or contractor.

This Handbook is based on engineering experiences and is primarily intended for use by 412th Test Engineering Group (TENG) flight test engineers and management. There was no attempt to "prove" via data analyses that a proposed method would work better than previous methods. In general that type of comparison would be impractical, but perhaps future efforts may be able to provide supporting data.

This Handbook should not be used as a “cookbook”. The intent is for flight test engineers on each project to contemplate each topic and decide how best to apply the lessons to their particular situation. For example, the people on each project should carefully consider the backgrounds of their counterparts and determine if it would be best to provide them with this handbook in its entirety… or not. In general, it might be best to discuss individual lessons during topic-related negotiations with the program office, contractor, or participating test organizations.

One of the main goals for documenting these lessons was to provide useful information to 412 TENG flight test engineers who become involved during the initial planning efforts prior to contract award. On many previous projects, flight test engineers did not get fully involved until long after contract award (sometimes years later). In those situations it was often not practical to implement many of these best practices because they were considered out-of-scope relative to the contract. That placed the government flight testers in the difficult role of advocating best practices that had to be funded above and beyond the original contract. Sometimes those efforts were successful, other times they were not (often to the detriment of the overall project). Fortunately, early tester involvement has become much more common and is now an imbedded part of the acquisition process. Unfortunately, large scale flight test envelope expansion projects have become much less frequent and it will be rare that individual discipline engineers have experience on multiple projects of that type.

USE OF THIS HANDBOOK

It is hoped that the lessons contained in this handbook will help guide future flight testers as they coordinate with other project stakeholders throughout the preparations for air vehicle envelope expansion. Each of those flight testers will have their own experiences that will influence what they attempt to achieve, but these lessons probably encompass more than the limited experience of any single individual. These lessons may also help provide some consistency for what flight testers attempt to achieve across multiple projects. With all of that in mind, the lessons are divided into timeline-oriented groupings to help introduce the concepts at an appropriate phase of the overall preparations:

- Topics to Coordinate Prior to Contract Award
- Topics to Coordinate Prior to First Flight

Those groupings are merely suggestions and flight testers will need to assess the requirements and timelines of their specific project. An intent of those groupings was to identify which topics would be best incorporated into contractual documents versus those that can wait to be coordinated as part of normal planning activities at a later date. There may be minor variations that would be warranted depending on the type of contract (EMD, Demonstration-Validation, etc.). Note that there is no grouping for "Topics to Coordinate after First Flight" because that would typically be too late to be effective for envelope expansion.

Each lesson learned is documented in the main text in a fairly standard Lesson Learned format as follows:

TITLE:

LESSON LEARNED: A few sentences describing the fundamental lesson.

PROBLEM: A description of the past problem(s) that the lesson attempts to avoid or improve.

DISCUSSION: Enough background to provide a basic understanding of the relevant history, technical aspects, probable tradeoffs, etc. This section is primarily intended for technical management but can also be used as a first look for discipline engineers.

APPROPRIATE ACTION: This section is probably the most important part of each lesson. It attempts to provide guidance to future flight test teams on how to advocate the recommended concepts (sometimes to a hostile audience, but more typically to an uninterested audience with their own priorities and philosophies). Many of the recommended advocacy techniques were successful on past projects. Some of the recommended techniques are suggestions that have not yet been attempted on flight test projects.

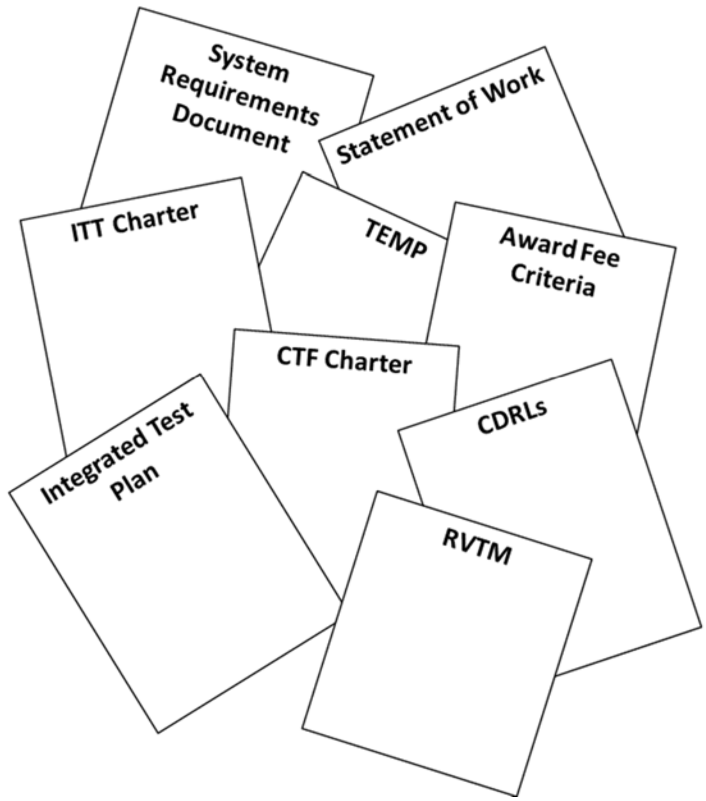
SUPPLEMENTAL DISCUSSION: This is a non-standard section that was added to dig deeper into the background needed to better understand a broader range of implications. This section tends to be oriented towards discipline engineers but should also be useful for technical management.

This lesson learned format is targeted at a flight test audience that would be most impacted by the negative aspects if the lessons from previous projects are not adequately addressed. There was no attempt to provide comprehensive technical background from specific historical projects to support the discussion of each lesson. That type of documentation would have required approvals from each and every controlling organization. Instead, the discussion in each lesson attempts to capture the essential concepts that have impacted multiple historical projects without focusing on any specific project. Some of the top-level technical aspects are discussed but if the user needs more explicit information from previous projects, they will need to accomplish the specific technical research necessary to effectively apply these lessons. In many cases that research will require approvals from the controlling organizations. Each lesson learned is also documented in Appendix A using a briefing slide format for easier presentation to audiences more interested in just the top-level concepts.

Supporting data may be available for some of these topics but generally the issues, impacts, and best practices are based on the experience of the individual engineers. Those engineers collaborated to identify topics that have occurred on more than one program to produce these lessons. In most cases, the customer on past projects would not have wanted to pay extra for the type of intensive tradeoff studies necessary to provide detailed data for future projects. Since it is not possible to "re-do" previous flight test projects with more of these best practices in place, the best way to obtain a broader evaluation of these concepts is to apply as many as practical to upcoming flight test projects.

These lessons are not envisioned as the final list. It is expected that this should be a living compilation that will continue to expand and evolve. New topics should be added and existing topics modified to stay current with ongoing trends in the acquisition and flight test process. It is recommended that any new topics remain consistent with the philosophy that they represent recurring themes on multiple projects, not the unique experiences associated with a single project. A few suggestions for "Possible Future Handbooks" and "Possible Future Topics" are in the Concluding Thoughts section.

TOPICS TO COORDINATE PRIOR TO CONTRACT AWARD



Anyone who is part of the USAF acquisition process knows that there are many established training courses that describe the overall process and how it is intended to operate. In spite of the abundance of training material, many people in the flight test portion of the acquisition process have felt “disconnected” from the contract development portion of the process. That has been especially true for the younger flight test engineers who were probably assigned to a given project many years after the contract award. In addition, the acquisition process is constantly changing and the contract development process described in training courses may lag behind actual practice.

However, one of the more important changes to the acquisition process has been to get flight testers involved much earlier as an integral part of the contracting process. This trend has provided the essential opportunities flight testers need to influence the contract in ways that can have very positive impacts on the ensuing flight test portion of the project. This does not mean that flight testers will have the ability to insert a standard “formula” into contract language. Any influence on the contract language will still need to fit within the overall vision for that project and be approved by the program office and accepted by the contractor(s). Therefore, flight testers who are asked to participate in the contracting process for future projects will need to adapt the lessons learned described in this handbook to accomplish the essential goals suggested for each topic by influencing contract language in ways that will stay consistent with the contract philosophy for that new project.

For most flight test engineers, providing inputs to contractual language will be well outside their “comfort zone.” However, more flight test engineers are being drawn into projects earlier and young flight testers may be surprised how soon that happens to them. It will be in their best interests to apply due diligence to this task. It is hoped that this handbook will be helpful to those individuals but there are also other sources of relevant knowledge. As an example, Appendix D provides a copy of an OSD Guide on Incorporating Test and Evaluation into DoD Contracts. That guide may be useful for overall concepts, and there may be more recent versions available.

There will be more than one option for how contractual language can be influenced by flight testers. If it is too late to influence the System Requirements Document or Statement of Work (SOW) for a given project, it may still be possible to influence the list of CDRLs, the Award Fee Criteria, the TEMP, the Integrated Test Team (ITT) Charter, or the Combined Test Force (CTF) Charter. In particular, the philosophy for how the ITT will operate will have a long lasting influence on the rest of the test program and can also have significant impacts on the reminder of the operational life of the platform. Therefore the concepts captured in the ITT Charter should be given careful consideration. In addition, the Integrated Test Plan (ITP) and Requirements Verification Traceability Matrix (RVTM) provide more opportunities to have a positive impact. Flight testers need to be involved with the development of both of these so that system-level test requirements are appropriately assigned to flight testing or other venues such as analysis or lab testing. The Integrated Test Plan tends to be at a fairly high level and should not be confused with a detailed flight test plan. The RVTM will be important because it establishes the basis for which specific test types are formally required to be accomplished in flight test. Well defined RVTM requirements can often be referenced in the opening paragraphs of a detailed flight test plan as clear justification for a given type of test. It is also important to understand that contractual language to capture flight test concepts does not need to be verbose. In fact, the shorter the better. Too much detail will just provide ammunition for those that want to object to the concept. The goal should be to clearly state the overall intent for a given topic in ways that are unambiguous but that are still flexible enough so that details can be worked out later.

Many of the recommended best practices would be very difficult to implement if the contractor and/or program office consider them “out-of-scope” per the formal contract. If the fundamental concept is properly addressed in the Statement of Work, it will be much easier to interpret a best practice as “in-scope”. As one example, it will be important to identify the overall requirement for Flight Test Aids that are applicable to the particular envelope expansion effort along with a statement of intent such as, “Flight Test Aids shall be sufficient to provide interim automatic limiting (e.g., g-limiting, roll rate limiting), control system excitation (such as a Flutter Excitation System or Pre-programmed Test Input), and make flight control system tuning adjustments during flight test”. There is no need to include a long, detailed list of Flight Test Aid mechanization options in contractual language since that list will probably need to change as the design matures.

As another example, it would be important to specify in contractual documents that a flight test support simulator at the test site is an expected capability, who will implement and update that simulator (contractor or government), at what point an initial capability is expected, and the overall concept for providing updates. There is no need to include a long, detailed list of simulator features in contractual language. It might be appropriate for contractual documents to include a broad description of the expected overall scope of flight test simulator functionality to help bound cost estimates, but without getting bogged down in detailed lists of features.

A golden opportunity occurs when flight testers are invited to be involved prior to the Request for Proposal (RFP) and the ensuing contract award process. Those flight testers will be given the chance to insert best practices and help avoid historically recurring problem areas. However, those flight testers will need to understand the situation they are entering. Much of the communication prior to contract award will be highly constrained by the competitive nature of the process. There will probably be two or more prime contractor teams all focused on winning the contract. Direct communications with contractor personnel will be VERY limited in order to avoid unintentional tainting of the process. While those limitations on direct communication will be necessary, it also makes it much harder to develop language for the key contractual documents that is acceptable to all participants. All proposals for the language in those key documents will need to be adjudicated though the Program Office. The Program Office may agree to accept inputs from flight testers but will then need to coordinate with each of the bidding contractors to identify common language satisfactory to all concerned. That process can cause the wording to “drift” away from the original intent. There may be a small number of iterations (if any) before the Program Office needs to freeze the language and release the RFP. The best flight testers can hope to accomplish is to capture the fundamental concepts and highlight to the Program Office any areas that appear to have “drifted” too much. Flight testers will need to take advantage of the few occasions where direct communication with the contractor is allowed (if any) to better understand the contractor point of view and attempt to help reword language as needed to meet the original intent while also avoiding problem areas from the combined perspective of all the bidding contractors.

If flight testers do not get involved until after the contract has been signed (the typical situation in the past), that does not mean the topics in the “Prior to Contract Award” category should be abandoned. In that situation, flight testers may need to get very creative on how to best communicate the benefits to other stakeholders. The primary option for advocating each best practice will be to emphasize the programmatic benefits in terms of the long term efficiency and effectiveness of the development effort. On recent projects, some of these lessons have been adopted with success, even after contract award. The existing contract language for a given project will establish the overall framework and boundaries for what can be accomplished. However, most contracts are intentionally written to be broad in scope and with a certain amount of flexibility. In addition, the contractual language may have enough ambiguity to enable best practices to be incorporated while staying within the scope of the contract or can be justified as a warranted risk reduction. Therefore even if an engineer is assigned to a project long after contract award, it will still be important for that person to be familiar with how the existing contract language influences each topic described in this handbook for their particular situation. That knowledge will enable the engineer to seek help from local management to incorporate as many of the best practices as can fit within the attainable boundaries for that project.

CA1: DON'T NEGLECT BASIC AIRCRAFT SYSTEMS

LESSON LEARNED: On large EMD-type test programs, the basic aircraft systems (i.e., propulsion, flight controls, flying qualities, structures, environmental controls, brakes, fuel system, hydraulics, basic communication/navigation [COMM/NAV], weapons bay acoustics, etc.) have had more developmental problems than expected. The first test aircraft produced should be capable of testing a broad spectrum of these basic aircraft systems

PROBLEM: It is understandable that new aircraft acquisition programs tend to focus on the advanced technologies required to implement game-changing capabilities (stealth, sensor-fusion, new weapons, etc.). Those are typically the capabilities with the highest priority to the user and with the greatest expectations from external oversight authorities (DOD and Congress). However, that emphasis has often led to neglect of the basic aircraft systems that were taken too much for granted. The worst cases have resulted when production representative basic aircraft systems were treated as secondary elements or not even implemented on the first few test aircraft (see examples in Supplemental Discussion). The end result has been long delays to envelope expansion testing which further cascaded into long delays to combat-representative testing of the high priority advanced technologies.

DISCUSSION: Basic aircraft systems are the cornerstone of the integrity and operability of the entire weapon system. However, they have typically been regarded as "routine" with less development risk than the more advanced systems to be integrated into the weapon system. This perception has tended to diminish emphasis and focus on their development relative to the advanced cutting edge technologies. Such perceptions have even led to implementing off-the-shelf systems in place of the intended production design in the early test aircraft as a perceived expediency. History has shown that delaying development of production representative basic aircraft systems often leads to serious consequences in the test program. Implementing non-representative systems and/or non-representative off-the-shelf "routine" systems in the early test aircraft for expediency has often had serious unexpected consequences that required significant amounts of time to resolve (see examples in Supplemental Discussion). This unplanned time was needed to keep the test program going with the problematic non-representative system, but did nothing to support the development of the production version of that system nor the overall weapon system.

A related scenario has occurred when off-the-shelf components have been included as an inherent part of the production design. Example: using landing gear from a previous aircraft. That may indeed be a perfectly reasonable design approach, but problems have occurred on multiple projects when far too much credit has been taken for the use of those off-the-shelf systems. The use of off-the-shelf components has led to false confidence that those systems will work flawlessly as part of the new design, and the normally routine testing of those systems has been neglected. The typical pattern starts with a less-than-rigorous review of the design limitations of the off-the-shelf system and insufficient attention paid to integration of that system into the new overall design. For example, the initial review of the landing gear from a previous aircraft may have confirmed that it is adequate to handle the full weight range of the new design. However, it is eventually discovered that the previous landing gear cannot handle some other critical requirement for the new system (such as sink rate, side loads, etc.). The intent of this example is not to focus on landing gear, but to encourage a more thorough review of all aspects of the new design and to conduct a more rigorous analysis and test process, versus the more typical series of assumptions that the old system will work fine when integrated into the new design. This may be more complicated when the old system has proprietary aspects that constrain a deeper understanding of the real limitations.

Programs need to recognize that the basic aircraft systems are the cornerstone of the overall system. Without the proper emphasis on these systems, the integrity and operability of the platform often suffers. The consequences to the efficient progress of the test program and weapon system development have been profound. Early flight testing should focus on these systems and expanding the flight envelope. Early testing of advanced capabilities could still be planned, but not at the expense of the basic aircraft systems. The first few test aircraft should be configured with the appropriate basic aircraft systems, structure, instrumentation, and flight test aids to focus on testing these systems to supplement the more traditional role of expanding the envelope.

APPROPRIATE ACTION: The commendable trend for early tester involvement in acquisition projects should be used as an opportunity to identify and rectify development plans that neglect basic aircraft systems. In particular, any development plan that implements non-representative off-the-shelf basic aircraft systems in early test aircraft should be challenged by advocating improved concepts at discipline-specific working group meetings, but progressing to higher levels if necessary. It is possible that project leaders may not be aware of the historical pattern of wasted development effort associated with non-representative systems. There is no guarantee that the test team will be successful in advocating improved development concepts (which are sometimes viewed as outside the purview of the test team). At the very least, project leaders need to be better informed of the historical consequences before final decisions are made.

On some future development projects it may become unavoidable to delay implementation of certain basic systems until after the first test aircraft. That could happen due to a wide variety of unanticipated events such as supplier problems, more precise analyses of the design, or failures during component testing. In those situations the only remaining option to preserve the overall schedule might be to implement an off-the-shelf system in lieu of the intended production system. This will lead to a mandatory retrofit later in the development cycle, along with the necessary regression or repeat flight testing. However, many previous projects have implemented non-production representative basic systems on the first test aircraft as a conscious decision based solely on perceived expediency, not as a last resort.

Another factor that can lead to implementation of non-production representative components in early test aircraft is the perceived imperative to control the Technology Readiness Level (TRL). Controlling TRL is often viewed as a risk reduction strategy in RFPs. Incorporating a temporary COTS solution that has demonstrated a high TRL on another aircraft (e.g., 8 or 9) may seem much less risky than using a new design with lower TRL (i.e., 5 or 7). However, that approach only works if a realistic level of rigor is applied towards the comprehensive integration of that COTS solution. It has been disturbingly common to base decisions on a philosophy something like, "It's COTS, of course it's going to work", and skimp on the comprehensive integration step during the design process. In fact, the more often phrases like "It's COTS, of course it's going to work" are heard, testers should start to ask more focused questions about what has been done to ensure adequate integration of that COTS solution.

The fundamental goal of this Lesson Learned is to encourage test teams to question those early decisions when there is still time to influence the choice. The best way to avoid conscious implementation of too many non-production representative basic systems is to specify expectations for the capabilities of the first test aircraft in the SOW or other contractual documents.

When the decision has already been made to use an off-the-shelf system as an inherent part of a new production design, flight testers need to encourage a more thorough review of the actual limitations of that system as compared to both the clearly defined and indirectly derived requirements for the new system. A good way to initiate this process is to coordinate with technical experts who have experienced problematic integration of similar off-the-shelf systems on other projects. Those technical experts often have specific historical examples where key characteristics of an off-the shelf system were overlooked and led to significant programmatic impacts. However, even when those historical examples have been communicated to designers, it has been common for those designers to focus on why those specific examples won't be a problem instead of expanding their review to envision other aspects that may actually turn out to be a real problem. This process may require some creativity on the part of both testers and designers to identify those problematic aspects BEFORE they are revealed during the normal course of flight testing.

SUPPLEMENTAL DISCUSSION: This Lesson Learned complements the Lesson Learned “CA2 EMD Test Aircraft Configuration and Utilization.”

The following items are generalized scenarios that have occurred on multiple new aircraft EMD efforts.

Communication is an essential function provided by a basic aircraft system. However, new aircraft tend to be designed with highly integrated communication systems that are often viewed as too complex to implement on the first few test aircraft. It has become typical to outfit the first few test aircraft with “off-the shelf” communication systems that are viewed as simple and reliable. Unfortunately, when those off-the-shelf systems are installed within a new aircraft structure significant and unforeseen problems have occurred with aspects such as electronic grounding, interference from other transmitters, etc. Many current and future USAF development programs include aircraft with low observable requirements which tends to drive an ever-growing percentage of composite structural elements (contributing to more difficult electronic grounding) and also constrains the ability to alter the outer mold line including apertures and antennas. These factors make it especially difficult to adequately integrate an off-the shelf, temporary communication solution. The result has made it extremely difficult for the test pilot to talk with the test conductor in the mission control room or with pilots in the chase aircraft or aerial refueling tanker. This has been especially disruptive during the early portions of major test programs when clear communications were a safety necessity during envelope expansion flight tests. It has taken many months to identify and resolve the issues, especially when no contingency plans were in place. All of that unplanned effort was applied to a temporary system and provided no improvements to the final product. The inability for the test pilot to communicate properly has directly contributed to significant programmatic delays right at the beginning of the flight test program. Ideally, the very first test aircraft would be outfitted with a production representative communication system. That approach may be more practical than initially perceived, as long as it is planned from very early in the development effort. If it truly is impractical to start flight testing with the production communication system, no assumptions should be made that an off-the-shelf system will work properly. Extra effort will need to be applied to ensure proper electronic grounding, adequate placement of antennas, etc. When in spite of best efforts the off-the-shelf system does not provide adequate functionality, quick-reaction contingency plans need to be in place. Appropriate personnel need to be contractually obligated and located at the test site to identify the root cause for the problems and to implement satisfactory fixes. Starting envelope expansion flight tests with a backup radio is a good idea but care needs to be taken that it doesn’t have the same problems as the off-the-shelf system. A poorly designed communication system can also have profound impacts during early envelope expansion of UAVs. Even though the UAV communication system does not impact pilot-to-test conductor conversations, it still needs to be reliable enough to make sure the new UAV test aircraft takes the right action at the right time.

Navigation is another essential function provided by a basic aircraft system. Modern aircraft typically use an INS integrated with a GPS. The early planners on new aircraft acquisition efforts often start with the reasonable-sounding assumption that the first few test aircraft will not need to penetrate deep into enemy airspace and will not need a production representative INS/GPS. An off-the-shelf INS/GPS was selected that was viewed as relatively simple and reliable to provide a source of that type of data for various purposes. Sometimes this off-the-shelf system has only been used for time-space-position information (TSPI) flight test data but many times the outputs were used as temporary inputs to other systems. Even when those off-the-shelf systems provided reasonably good information about aircraft position, intermittent problems existed that were disruptive to envelope expansion progress. The root cause of those problems needed to be identified and resolved before efficient progress could be maintained. It was common that the subcontractor personnel needed to accomplish those tasks were not even on contract because the assumption had been made that the off-the-shelf system would perform flawlessly. It was also common that some of the problems were due to integration issues that were overlooked during a design process that had been abbreviated because it wasn’t a production system. None of the effort required to investigate and repair these problems was applicable to the production navigation system. Serious consideration should be given to outfitting the very first test aircraft with a production representative navigation system. That approach may be more practical than initially perceived, as long as it is planned from very early in the development effort. If it truly is impractical to start flight testing with the production navigation system, no assumptions should be made that an off-the-shelf system will work properly. Appropriate contingency plans need to be in place.

Weapon Bay Doors may not be considered a basic aircraft system; however, they are becoming a ubiquitous component on many aircraft with internal weapon requirements. The problems have begun when the third test aircraft off the assembly line (or later) was designated as the primary weapon separation platform. That designation has changed the mindset of the entire design team to reflect the philosophy, “We don’t need any production representative weapon bay capabilities until test aircraft number 3.” The result has been that many components associated with the weapon bays were not delivered with the first two test aircraft, or were replaced by non-representative components in a narrowly defined effort to save money or schedule. The hidden costs to that approach were not discovered for many years and those costs could easily be an order of magnitude higher than the original savings. As an example… the first two test aircraft have been delivered with non-production weapon bay door drive mechanisms (including the actuators, electronic components, and computer control elements). That approach allowed the weapon bay doors to be opened in early flight test for limited testing. However, the non-production components did not allow the doors to be opened at the same rate as would be achieved on the final design (usually slower than the final design). Therefore any aircraft response transients associated with the weapon bays opening and closing would not reflect the actual transients that would occur with the final design. The end result was that realistic transients were not discovered until testing on test aircraft number 3. When it was necessary to significantly modify numerous components (not limited to the flight control laws) to address transients that were greater than expected, the planned weapon separation process was delayed for many months. Similarly, production-representative weapon bay door opening and closing rates, which may consist of very quickly opening and closing the doors so as to meet operational needs, may have significant effects on aircraft structure and components. If early tests use different opening/closing rates and components, the true structural effects of weapon bay door operation on the production configuration may not be known early enough in the test program to be corrected without significant schedule and cost impacts. Since the end-to-end weapon separation process is always one of the longest and most critical paths during development, this can be a major contributor to delaying the Initial Operational Capability of the entire weapon system. In the long run it would have been far more cost effective to implement representative weapon bay door actuation components much earlier, discovered the problems, and designed fixes. The lengthy weapon separation process could then have proceeded much closer to the original schedule. Similar hidden costs have been incurred when the first test aircraft was not fully instrumented for weapon bay environmental testing such as vibroacoustics and internal loads (also because the mindset was that those capabilities would not be needed until test aircraft number 3). Instead of delaying weapon bay open environmental testing until a later test aircraft, it would be far better to incorporate that test path as a normal part of envelope expansion. That type of testing probably doesn’t need to be part of the leading edge (i.e. along with basic flutter testing), but should follow as quickly behind as practical for that project. In summary… even though the first test aircraft may not need to drop or release weapons, it needs to have enough fundamental weapon bay related flight test capabilities to allow for effective envelope expansion testing. This aspect is related to similar topics in the Lesson Learned “FF2: Flight Test Flexibility”.

Structure is an essential basic aircraft system. It is the foundation of the entire airframe and defines the flight limits, maneuverability, and load carrying capacity. On major EMD type programs the structure of the first test aircraft typically is not production representative. This was usually due to negative structural margins that were revealed by more precise analyses as the design matured, but those negative margins were discovered too late to redesign the structure of that first test aircraft. This concept is discussed more explicitly in the next Lesson Learned “CA2: EMD Test Aircraft Configuration and Utilization.”

CA2: EMD TEST AIRCRAFT CONFIGURATION AND UTILIZATION

LESSON LEARNED: On EMD-type test programs, the first test aircraft was often under-utilized and unsuitably instrumented because it was envisioned for an artificially narrow scope of test types. The end result has been inefficient use of resources, delays in other types of testing, and excessive flight time testing non-representative configurations.

PROBLEM: On large EMD programs, the contractor has often been incentivized to meet delivery dates for the test aircraft. As the design gets behind schedule, which has often been the case, decision makers sacrifice capabilities in the first few aircraft to meet the incentivized dates. Therefore, the first test aircraft is usually not very production representative. The first aircraft usually has significant flight limitations, yet the aircraft is likely extensively instrumented for flight testing across the entire flight envelope with the assumption that it will be the "workhorse" for that type of flight testing. Even when parts of the configuration are not production representative, accomplishing preliminary testing of that non-representative configuration often becomes a high priority in order to show some kind of progress regardless of the inefficiency. It is also common that the aircraft is not configured or instrumented for other test types that could have been conducted on that airframe even within its unplanned limitations. The overall result has been:

- Less capable test aircraft
- Later delivery of the first test article due to installation of extensive instrumentation during construction (even though that instrumentation is of little use given the limited envelope)
- Flight time spent to test non-representative configurations producing data of little relevance to the production configuration
- Underutilization of an expensive test resource
- Commensurate increases in the cost of the test program.

DISCUSSION: The history of large EMD programs shows that the first test aircraft was usually not production representative with respect to key capabilities, even though allocation of flight test tasks presumed those capabilities would be production representative. The following examples provide some of the consequences experienced when this likely state of the first test aircraft is not recognized in planning.

One major example that has been a historical constant was that the structure of the first aircraft was not production representative. The structural design to meet the combination of strength and performance-related weight goals was typically behind schedule, so the structure of the first test aircraft was compromised to meet the delivery date, resulting in significant flight limitations. However, in the early planning stages, this historically repetitive fact was overlooked, and the first aircraft was usually extensively instrumented for loads and flutter flight testing during construction. The loads and flutter instrumentation planned for that first aircraft was extensive, requiring installation as the plane was being built, adding significant time and delaying first flight. Effective loads and flutter flight testing require the aircraft to fly to the full design envelope of the airplane, which the first airframe was limited from achieving since it wasn't structurally representative. Therefore, very little effective loads and flutter testing has been achieved on the first test aircraft and it was necessary to repeat much of the limited flight testing on a later test aircraft.

On a recent major project, the first test aircraft was fully instrumented for loads testing, but it was recognized long before first flight that the structure would not be production representative and that the envelope expansion capabilities of that aircraft would be very limited. The decision was made to cancel the planned calibration of the loads instrumentation, so at least the project saved the 1-2 months of time and related expense. How much more time and money would have been saved by instrumenting that first test aircraft with only enough loads instrumentation to clear a partial envelope, and plan for full envelope clearance on a different, more representative test aircraft? The essence of this Lesson Learned is that this pattern has been normal on many projects, not a unique aberration for that specific project. In fact, recognition of that pattern has not been limited to recent projects but was also observed many decades ago and was reflected in MIL-STD-1530 in 1972 as described a few paragraphs below.

It may be possible to achieve significant programmatic efficiencies by recognizing the historical facts and consciously installing less loads and flutter instrumentation on the first test aircraft. It is understood that this is a radical concept that does not fit well within long-accepted principles that have been applied on numerous large scale projects (but with highly questionable results). Clearly, there would still need to be enough loads and flutter instrumentation to support envelope clearance of a useful partial envelope. The partial envelope would need to be defined in a way that makes sense for a given project. An aircraft with a very small final envelope (such as a low-maneuverability unmanned aerial vehicle [UAV]) may not see as much benefit from this alternative approach as an aircraft with a very large, multi-dimensional final envelope (such as a supersonic, high-maneuverability fighter). The limited loads and flutter instrumentation would need to be sufficient for fundamental safety assessments (e.g., control surface loads and accelerometers, etc.), but would not need to include the full set of strain gages, accelerometers, etc. needed for full model validation. It may even be possible to avoid installing and calibrating instrumentation on the first test aircraft that would typically be viewed as “essential” such as wing root bending and torsion. As long as the resulting envelope is useful for other purposes, the “payoff” in savings would be highest by avoiding the need to install the first aircraft into a complex calibration fixture. On the other hand control surfaces loads are usually instrumented via actuator and are relatively easily calibrated. The end result would be a test aircraft that could not clear structures across the full envelope (which has been the historical situation), but that could get to first flight sooner given much less time for installation and calibration of extensive instrumentation. The limited set of instrumentation would still be useful to verify the methodology and get an initial look at model correlation.

The structure of the first test aircraft would still need to be designed with sufficient margin so that the initial envelope is large enough for other disciplines to accomplish useful flight test. Recent proposals for new aircraft have emphasized structural test programs that provide for more margin on the initial test aircraft and therefore result in greater limits on those aircraft with less reliance on in-flight testing. It is hoped that this trend continues. The instrumentation for the other disciplines on the first test aircraft would be much simpler than loads and flutter instrumentation. Examples include flying qualities, aerial refueling, propulsion, subsystems, and fundamental avionics and mission systems (advanced avionics and mission systems may not be necessary).

The primary instrumentation for structural flight testing across the full envelope should be planned for a later test aircraft. The second test aircraft off the assembly line is typically too close behind the first test aircraft to implement the required design changes without excessively disrupting the assembly line. Therefore, it would be more prudent to plan on the third test aircraft (or later) to be built with the primary structures instrumentation installed. This would provide designers extra time to fine-tune analyses and update the structural design appropriately and would provide flight testers with a suitably instrumented test aircraft with more production representative structure. The genesis of this concept goes back circa 1972 in MIL-STD-1530A, *Military Structural Integrity Program* (reference 1). The more recent wording from MIL-STD-1530C, *Department of Defense Standard Practice: Aircraft Structural Integrity Program* (reference 2) is: "An additional aircraft, sufficiently late in the production program to ensure obtainment of the final configuration, shall be the backup aircraft for these flight tests and shall be instrumented similarly to the primary test aircraft." On almost every major EMD flight test effort over the last three decades it has been a near-universal truth that the “backup” structural test aircraft became the primary because the structure on the first test aircraft was not production representative.

Another historical example that has been consistent in EMD programs that include low observables (LO) is the LO configuration of the first test aircraft has not been production representative. As with the previous structures example, the LO design was behind schedule so the LO configuration of the first test aircraft was compromised to meet the incentivized delivery date. However, program planning usually included accomplishing preliminary Radar Cross Section (RCS) testing on this first test aircraft as a high program priority to show "progress" early in the program. These preliminary RCS tests became high priority work for the first test aircraft, but with the likelihood of the results having little relevance to the production configuration. In addition, the very limited flight envelope of the first test aircraft has inhibited effective and efficient LO flight testing. The strict LO configuration maintenance took significant down time on the airframe and may even have delayed first flight. If the historical facts that have limited effective LO testing were acknowledged, LO would not be planned as a major capability of the first test aircraft. The aircraft could be configured for other testing and delivered sooner. It is acknowledged that this concept may be very difficult to communicate to people and organizations that are very highly focused on getting LO flight test "answers" at the earliest possible opportunity. Since the situation described has occurred on many EMD programs with LO test requirements, the end result should provide compelling evidence of its likelihood. The decision makers on every test program will need to assess the likely tradeoffs between essential need and disruptive effects, but it is hoped that this lesson will inspire a more realistic assessment.

Typical planning by contractors for the utilization and configuration of test aircraft has historically reflected a tendency to label individual test aircraft with narrowly focused purposes. The rationale has been that they perceive it as a primary way to control costs. They optimistically plan to complete air vehicle testing very early in the program, which then releases workforce and resources and allows the focus to shift to mission systems. History has shown that this has not often been achieved. Air vehicle testing was slowed and delayed by the compromises in the designs to meet aircraft delivery dates, while opportunities for earlier looks at mission systems were missed. A significant portion of all EMD testing could be accomplished within the limited envelope of the first few aircraft. Much of the avionics testing could be accomplished on these airframes with the right instrumentation and some representative elements of the mission systems (i.e., radios, mission planning, etc.). Early versions of the avionics have usually already been through the system integration labs and flying test beds. Having looks at these early versions on the actual platform provides earlier insights into potential issues.

A potential contractor concern with the alternative test aircraft utilization as described in this Lesson Learned might be the implications of incremental envelope expansion and early mission systems testing on control room personnel. Contractors tend to plan large flight test programs with the assumption that mission systems types of testing will not need to be accomplished until the envelope has been fully expanded and therefore large numbers of air vehicle-oriented engineers will no longer be required in the control room (thereby resulting in significant cost savings). That approach may seem very logical and cost effective, but it has hardly ever worked out that way on large EMD projects. For all the reasons listed in multiple lessons learned, the envelope was often not fully cleared when mission systems testing needed to begin. This unplanned situation has forced air vehicle-oriented personnel to remain in the control room far longer than intended, resulting in more expensive contractor operations (or a shift in workload to government personnel). Although the government cannot direct the contractor on how to staff flight test operations, it may be warranted to suggest a more balanced alternative in which a smaller cadre of air vehicle and mission systems personnel are kept over a longer time period (versus the more typical plan to staff very heavily in a given discipline, finish a task, and then eliminate the personnel no longer considered necessary).

Weapon separation testing is typically a "long pole" in the testing schedule because of data turn-around times and analyses required between drops or launches. Accomplishing as much weapon separation work on the first aircraft as practical early in the program would reduce the workload later in the program where there is typically an overwhelming bow-wave of work. If the contractor incentives were based on delivering capabilities in the first few test aircraft that recognized the historical pattern from previous EMD test programs, test execution and progress would be much more effective and efficient. That overall efficiency could more than make up for the perceived cost efficiency due to optimistic workforce reductions. Capturing these concepts within the contractor incentives has indeed been happening on recent projects with improved results. In addition, the processing times for weapon separation testing have been much reduced, but should still be considered a "long pole" for aircraft with many bomb and missile configurations.

APPROPRIATE ACTION: When planning the test aircraft configuration and utilization for large EMD programs, the contractor should be incentivized to emphasize delivery of critical capabilities for the first few test aircraft, not just a delivery date. The planned configuration and utilization should reflect that design delays are inevitable, leading to compromises in the design of the first few aircraft, and quite likely significant flight limitations. Narrowly focused purposes for the first few test aircraft should be avoided, planning instead for a broader general utility of these aircraft in conducting flying qualities, aerial refueling, propulsion, subsystems and some weapon separation testing, even in the limited envelope. Since much of the mission systems work can be accomplished in a limited envelope, some mission systems capability should also be considered for these first few aircraft, getting an earlier look at those capabilities.

SUPPLEMENTAL DISCUSSION: The following sequence provides a step-by-step description of the process that has been observed on multiple large scale EMD development projects. Observing this process repeatedly was one of the main motivations for capturing this Lesson Learned.

Impact of typical structural design process

- EXTREME emphasis on weight reduction to meet performance requirements
- Leads to optimistic interpretation of structural margins
- Later, more rigorous analyses reveal dozens or hundreds of negative margins
- Many negative margins not fixable prior to first flight
- Each negative margin requires some type of flight limitation
- Some problems fixable with retrofits and lengthy downtime
- Many problems can only be fixed on later aircraft
- End result is permanent limitations on first aircraft
- First aircraft planned and instrumented as a structural test "workhorse"
- Much of the instrumentation and work must be shifted to a later aircraft
- Ship 2 often too close behind Ship 1 to implement fixes
- Sometimes Ship 3 can be built with production representative structure
- The original performance requirements that forced the extreme weight reductions were no longer considered important. The general view became< "The aircraft is what it is and will not be redesigned to meet those performance goals".
- Ship 1 retires early as: maintenance trainer, live fire target, decoration, or museum piece

CA3: EXPECTED TEST EFFICIENCY FOR TEST PROGRAM PLANNING

LESSON LEARNED: Planning for large flight test programs has used optimistic flight test efficiency factors.

PROBLEM: Planning for large flight test programs typically use flight test efficiency factors greater than historical statistical efficiencies from previous similar programs. The actual flight test efficiencies are much less than planned, leading to protracted test schedules, resultant higher costs than planned, and the appearance of poor program progress.

DISCUSSION: Flight test efficiencies on large EMD projects have historically averaged 60 to 65 percent over the course of the program. That is, 60 to 65 percent of the flying time is spent in the actual execution of test points. The remaining 35 to 40 percent of flying time has been needed for basic aviation activities such as takeoff, transition to test area, maneuvering to stay within test area, aerial refueling, return to base and landing, etc. Early in the program, the efficiencies may be lower, but the efficiencies gradually increase to average about 60 to 65 percent. Many large flight test programs have used flight test efficiency planning factors much higher (80 percent or more) during the planning stages of the program, to include an underestimation of the number of test points that need to be repeated over the course of the program (which has exceeded 50 percent on several projects). This leads to estimates of flight test hours much lower than actually required, which results in shorter schedules and less cost than will actually be incurred.

Convincing people to use realistic flight test efficiency numbers in their forecasts can be an uphill battle. Many new projects begin with the implicit expectation that a new process or philosophy will enable them to avoid the problems that have plagued numerous previous projects. Contractors have often had no choice except to base overall estimates on optimistic efficiency assumptions in order to meet customer expectations (or risk losing the contract). In an ideal world, the source selection process would at least compare the efficiency estimates of each contractor to try and determine which estimates are more realistic. There may even be previous examples that support high flight test efficiency values, but those have often been based on prototype or demonstration projects that didn't actually need to result in delivery of a final product.

APPROPRIATE ACTION: Programmatic planning should include experienced test organizations to provide recent examples of relevant flight test efficiencies (similar size program, similar type aircraft) for use in determining the amount of flying time needed to execute the program, and the corresponding schedule. These efficiencies should be the basis for all flight test planning schedule development for the program.

Key government personnel on various historical projects have kept their own versions of that type of data for the projects they worked on. If engineers working on new projects could communicate with those key individuals, they could often get copies to help guide their own forecasts. However, if those key individuals had retired or transferred, it has been very surprising that there has been no effective central government database where that type of fundamental data could be obtained (except possibly the base history office… which might not keep records for the particular data type of interest). Contractors have been much better at keeping very detailed records for each of their projects because they know that kind of data might help them win future contracts. However, they typically treat the raw data as highly proprietary and keep it closely guarded. In addition, the specific historical records being presented may be slanted towards an optimistic or incomplete interpretation as discussed in "CA8: Tracking Flight Test Progress." It is recommended that a government flight test organization be selected (and funded) to maintain that type of database for past projects, ideally using reasonably consistent methodology.

SUPPLEMENTAL DISCUSSION: The best results have been achieved when realistic flight test efficiencies were built into a project from the earliest scope estimates and maintained throughout the duration. The main impact of this approach is to highlight when the ”stretch” of a project has exceeded its “grasp” and can result in early scope reductions that help make the project more achievable. Once a given project has committed to an unrealistic scope, it may be politically too late to inject more realistic planning factors.

Obtaining detailed historical data from previous similar projects may be helpful to establish a solid baseline, but it is also likely that those projects will be viewed as “poorly managed” or otherwise not applicable. Going through a typical test mission from those previous projects may be helpful to focus attention on just how hard it is to come up with radical improvements. A given type of test aircraft will always take a certain amount of time to climb up to test altitude or transition to the test area. Takeoff and landing cannot be hurried. Aerial refueling can provide great overall efficiency improvements compared to refueling on the ground, but it still takes a certain amount of flight time to offload a full tank of gas. It may be practical to accomplish some types of test points “piggyback” with those other activities, but that only works when ALL of the details specified by the test point are achievable on the day of test (unlikely). Every project has developed its own techniques for managing “piggyback” test points. Some projects have differentiated between truly “piggyback” test points and “concurrent” test points. For truly piggyback test points, the necessary data can still be obtained even if the flight conditions for the “parent” test point change… or can be obtained in another way if the “parent” point is deleted). Concurrent test points have been treated as independent test requirements that have been selected at the similar conditions to try to improve overall efficiency. If one of the concurrent test points is changed or deleted (for whatever reason) the other point would remain. The problems with this approach tend to occur when people want to take credit ahead of time for the implied efficiency of concurrent test points, when it isn’t yet known if those points will actually be accomplished at the same time or not.

On many projects, the best that most flight testers can hope for is to move unrealistic planning factors in the right direction. If the planning factor is moved from 80 to 70 percent (instead of the more historically justifiable 60 percent), at least the project planning will be somewhat more realistic. It may also be advisable to document and save any historical data that was used to try and persuade decision makers. That data may come in very handy years later when a different group of decision makers is trying to understand why things haven’t worked out as expected.

CA4: IMPLEMENT LOW COST FLIGHT TEST SUPPORT SIMULATOR

LESSON LEARNED: Availability of a local flight test support simulator has enabled test teams to improve flight test safety, efficiency, and effectiveness, for a relatively low cost. This lesson is focused on envelope expansion testing for Air Vehicle disciplines, but the generic lesson also applies to Mission Systems disciplines.

PROBLEM: Without a local flight test support simulator, the test team was forced to prepare for test missions by traveling to a remote location to use the system development simulator, typically at the contractor's main site. On some projects that has resulted in freeway round trips of 4 hours or more. On other projects, getting to the only available simulator has required air travel and two or more days away from the flight test location. The end result has been that those trips were more infrequent than appropriate and the test team was less prepared for the specific situations expected on upcoming missions. The travel and time away from the test site added significant costs and schedule delays. Scheduling problems have also arisen when the contractor's simulator was also needed for ongoing, time sensitive development work. When that simulator was required for flight test support, it was often necessary for the development work to be interrupted and the simulator reconfigured to the current flight test version of the system, creating delays in the system development schedule and potential for configuration control and inconsistency issues. Some contractors have multi-purpose capabilities in their simulators enabling support of multiple projects… but this adds to scheduling problems, especially when the project needing simulator time is considered lower priority than other projects. In addition, it is common practice to require that mission rehearsals for high risk tests be accomplished within a few days of the planned mission, typically with the same aircrew. In that situation it is very disruptive for the required people to cease activities at the test site and travel to the distant, hard-to-schedule simulator.

DISCUSSION: Simulation has been a significant factor in improving flight test efficiency and safety. Many test points are difficult to achieve without practice and a means to establish the proper entry conditions that will result in the right combination of parameters needed for the test (e.g., loads, stability and control, high angle of attack). A flight test support simulator provides the means to accomplish such practice and experimentation to find the right set up conditions during test plan development and test card generation. A local flight test support simulator is also needed to accomplish mission rehearsals immediately prior to the flight test with the specific pilot, test conductor, and discipline engineers. The concept for this type of dedicated flight test support simulator is not new but can be traced to 1991 and earlier (*Configuration for Improved Flying Qualities Testing*, AFFTC-TIH-91-01, reference 3).

Some tests have much more associated risk and require a build-up to achieve the end points. The simulator provides a means to predict the results throughout the build-up sequence, to determine if the actual characteristics of the aircraft are matching the predictions during build-up test points. This improves the ability of the test team to develop Flight Test Continuation Criteria (FTCC) well before the test mission. The FTCCs provide the test team with consistent ground rules to use to decide whether to continue on the intended path or to stop before reaching an unsafe condition if the actual characteristics are trending worse than the predictions (i.e., stability and control, high angle of attack, flight control system testing). Having the ability to do such simulation in preparation of the test plan, but also just prior to a flight provides significant improvement in the efficiency and safety of the actual flight. Numerous in-flight repeats are precluded which are very costly in wasted flight time, and the risk of potentially serious incidents to include the loss of the aircraft and life are greatly reduced. In addition, if unexpected results occur during a flight test, the readily accessible simulator (i.e. without excessive scheduling restrictions) enables the team to assess what occurred and can provide a quick return to flight test. Every test program that has employed a local flight test support simulator at the test site has experienced all these benefits and more.

Flight test support simulators have sometimes been rejected due to perceived cost. Simulators in general are often perceived as very expensive because simulators used for detailed design work typically include features like: full dome projection systems, flight hardware for key components (flight control computers, etc.), fully functional cockpits, and complicated motion base systems. Flight test support simulators for envelope expansion do not need any of those features. An “iron bird” simulator (that includes hardware components for the actuators and hydraulic systems) is normally well beyond the capability needed at the flight test location. In addition, existing generic simulators are usually available at some flight test facilities that can easily be configured for the aircraft type (e.g., fighter or heavy).

Flight test support simulators can be sufficient with software models only, but require hi-fidelity software models of the flight controls, aerodynamics, and propulsion systems. However those models almost always exist as part of the normal contractor development process. The effort to re-host those models into the flight test support simulator has not been excessive. It is important that those software models be updated to stay consistent with the test aircraft, but those updates almost always existed and just needed to be available to the test team. The best results have been obtained when the models were provided as source code. This allowed the test team to quickly react to specific flight test needs without the requirement to renegotiate support from the design team. For example: contractors are typically obligated to provide modeling capability for 2 or 3 off-standard-day atmospheric conditions (e.g., Hot or Cold). But flight testing is often conducted with very hot conditions at lower altitudes combined with an inversion layer that results in standard temperatures at higher altitudes. Source code has enabled the test team to insert the forecast temperature profile into all the correct locations within the atmospheric and propulsion models. The resulting predictions provide much more representative aircraft performance for a given test point at the forecast test day conditions, with a direct impact on determination of setup conditions.

Use of models provided by the designers avoids many credibility issues that would arise with use of independently developed models. Examples of capabilities that are often an inherent part of developmental simulators, but also necessary for flight test simulators includes features like: realistic gust inputs, crosswinds, aerodynamic parameter variation, flexibility in establishing mass properties (lateral CG in addition to more typical longitudinal CG, inertia characteristics equivalent to all achievable fuel states or internal/external store locations), etc.

One of the larger expenses for a flight test support simulator has been the primary controls for the pilot (stick/yoke and throttles). Those controls need to have tactile feel and functionality similar to the test aircraft, but they do not need to be flight qualified hardware. The cost for those controls can be minimized by using prototype hardware or by having units replicated to appropriate specifications by a low cost vendor. The remaining cockpit interfaces do not need to duplicate flight hardware but should be in approximately the correct location and provide a reasonable replication of the functionality. This can be accomplished with the least expense using reconfigurable touch screen displays. The “out-the-window” view for a flight test support simulator does not need to use state-of-the-art projection systems, but can be assembled using relatively inexpensive hi-definition flat panel televisions. Existing generic simulators at some flight test facilities may already provide these or equally useful features.

Any flight test support simulator would be most effective if combined with sufficient monitoring displays to support a basic control room team of Test Conductor and key discipline engineers. Ideally those displays would be identical to those used during actual test missions. The intent can be met with a small control room situated next to the simulator, or by connecting the simulator to the actual control rooms used for test missions. These capabilities are available with the generic simulators at some flight test locations. Some projects may feel that a small control room located next to the local simulator is not necessary if a connection to the main control room is provided. On large flight test projects using shared, centralized control rooms, experience has shown that the main control room is only used for a few mission rehearsals early in the test program and for occasional refresher training for the entire control room team. The extra cost per session for using the main control room tends to prohibit using it for normal mission rehearsals. Using the main, shared control room also incurs significant scheduling conflicts with ongoing flight test missions. Some flight test projects have had the luxury of a control room dedicated to that project/test aircraft. In situations where there were only one or two test aircraft sharing that control room resource, it was quite normal and effective to accomplish mission rehearsals using the dedicated control room while connected to the flight test simulator. This was practical because deconfliction with ongoing flight testing was completely under the control of that project.

Normal mission rehearsals can be supported by a small team of essential participants and do not require the entire control room team to take time away from other tasks. Some projects have located the small simulator control room in a separate building than the cockpit or on a different level within the same building. This layout has been impractical when the pilots need more guidance on sim setup or maneuver execution than can be effectively conveyed via a com system. It has been most practical when the small control room is located reasonably close to the cockpit.

Many projects have experienced a substantial exodus of qualified Test Conductor's and Discipline Engineers right in the middle of envelope expansion (often due to nothing more than the unfortunate timing of normal Permanent Change of Station for military people, and transfers to other projects for civilians). In those situations it can be very tricky to make the transition from a highly experienced and cohesive test team to a group of people who are essentially rookies… at least for that project. The flight test pace for that project is not allowed to slow down just because the key people have moved on. The availability of a local flight test support simulator has allowed those projects to withstand the massive changeover of personnel without excessively negative consequences. The new Test Conductors and Discipline Engineers were able to use the local simulator (combined with adequate control room functionality) to achieve a working familiarity with the design of the test aircraft and the flight test concepts specific to that project.

Fortunately, many of the topics described above have become much more common on new projects and the acceptance of those concepts is generally much higher than in the past. The best results have been obtained when the fundamental requirement for a flight test simulator is clearly documented in the RFP.

APPROPRIATE ACTION: It is very important for envelope expansion projects to have easy access to a local flight test support simulator and projects should not get hung up about who will operate that simulator (contractor or government). Both contractor and government simulator personal have demonstrated excellence when supporting flight test projects. Contractor personnel tend to provide more in-depth knowledge of simulation models and have better communication paths to designers. Government personnel tend to be more flexible when unforeseen simulator development needs to take place (i.e., outside the scope of a contract that was probably established years before the actual need). Regardless of who operates the simulator, the best models will be obtained from the contractor and that support needs to be part of the contract. That support needs to specify keeping the simulator in a useful configuration that is representative of the test aircraft including updates, capability improvements, hardware and software support, etc.

Each new flight test project should aggressively pursue implementation of a local, low cost flight test support simulator. Most major flight test facilities have generic flight simulators intended to support flight test. This process should include the appropriate agreements with the contractor to provide the necessary models and keep them updated. Ideally, the simulator would be located at the main test facility, but economies may be achieved by using generic simulation facilities within a short driving distance. Each project will need to assess the cost-capability tradeoffs. However it will be more important for each project to have a useable local flight test support simulator with minimal capability than to have none at all.

Test teams should be wary of any proposal to deliver a local flight test support simulator that is "just like the operational trainer". That may seem like a very logical and cost effective solution to distant management, but they don't need to actually do the flight test job. The problems with that approach may not be apparent until well into flight test. Operational trainers are typically designed for different purposes and do not include essential flight test-oriented capabilities such as: flight test unique cockpit displays, flight test aids, off-standard day temperatures, etc. Operational training simulators have typically used excessively simplified models and often do not include basic and vital flight test support features such as data recording. Also… operational trainers are typically not fully developed until long after envelope expansion. The constraints that tend to come with operational trainers can cause highly disruptive impacts on flight testers. There may be some benefits of using the flight controls and the cockpit "shell" provided with an operational trainer if that doesn't drive costs excessively. There may also be some benefits associated with the Emergency Procedures capability often provided with some operational trainers… as long as those procedures also reflect the current state of the flight test aircraft (but this is typically NOT the case). Using an operational trainer that does not reflect the basic Emergency Procedures needed during flight testing can end up providing the test team with highly negative training.

The best results for a local flight test support simulator will be obtained when the core of the functional design is based on the contractor's developmental simulators. In fact, one of the most cost effective approaches can be to replicate the physical and functional capabilities from the contractor's developmental simulators (at least in part)… as determined by the test team and program office. The cost effectiveness occurs because the flight test simulator will be based on a proven design using existing components, without incurring extra costs for developing brand new equipment and by avoiding expensive components such as motion base systems or fancy visual systems. The flight test support simulator probably will not need EVERY capability from the developmental simulators. The key overall concepts should be captured in contractual documents such as the SOW. The specifics can be negotiated later between the test team, sim operators, and management.

A flight test support simulator will be most effective if it includes certain key capabilities, even if those capabilities don't exist in the contractor's developmental simulator. Some examples like the ability to adjust longitudinal and lateral center of gravity (CG) will probably be inherent to the developmental sim. However, the developmental sim may be limited to a standard fuel burn curve or have other constraints whereas the flight testers will need to test the entire CG range established by flight test plans. Another example is the ability to adjust control surface effectiveness or other aerodynamic parameters. That capability will better enable the test team to recognize and prepare for unpredicted aerodynamics. One successful application of this capability has been to double the control surface power in any axis (equivalent to a 6 dB gain margin). Even though the test aircraft aerodynamics may not be expected to be off by that much, the ability to confirm that much margin in the flight test simulator provides test team personnel and safety reviewers with much more confidence. It also provides the test team with insight into the expected characteristics if the aerodynamics are only partially mispredicted.

SUPPLEMENTAL DISCUSSION: A concern often expressed is that it will become a requirement that ALL test points be rehearsed in the local simulator for EVERY flight test mission. Extreme requirements like that should be avoided. Abuses have been experienced in both directions. Some projects have had access to a local simulator but chosen to use it infrequently. The fallacy of that approach has been revealed during mishap investigations when it became clear that the mishap was predictable if the simulator had been used appropriately. Some projects have gone overboard requiring a repeat rehearsal if any control room member changes for a given mission or if only a few days have elapsed since the last rehearsal. Each project needs to define the right balance. A few general concepts for flight test simulator utilization are:

- All envelope expansion test points should be accomplished in a simulator at least once prior to attempting those test points in flight. This can be done at the contractor's main site, at the test site, or a combination. If some of those test points were accomplished long before the test planning process was completed, it may be appropriate to repeat some of those points to verify that the techniques and conditions are actually feasible and will get the desired data. The data from those sessions should be saved and readily accessible to relevant test team members.

- Simulator mission rehearsals with the full control room team are warranted early in an envelope expansion test project. The main intent is to get all test team members (often with a wide variety of past experience) to function as a unified team. This is a good time to expose the entire team to a wide variety of Emergency Procedures (not just in their own disciplines) to help each individual understand how they need to fit within the larger team during those urgent situations with interrelated and well-timed actions needed from each participant. This is also a good time to expose the team to key examples of what can happen if the procedures are not followed precisely as written in the test cards or if the systems do not function exactly as predicted. This helps prepare the team for surprises instead of just showing the team repetitive examples of everything working fine.
- Normal mission rehearsals can be accomplished with a partial crew of representatives from the most critical positions for the upcoming mission. This does not require every person from every discipline. A logical quorum that has been used effectively includes: Test Conductor, Primary Pilot (for multi-crew aircraft), and at least one engineer for each critical discipline.
- Normal mission rehearsals do not need to accomplish every test point planned for that mission. The focus needs to be on the most difficult test setups, leading edge envelope expansion test points, and any points with greater uncertainty in the expected results. Test points considered “routine” do not need to be rehearsed for every mission. Each project will need to determine where to draw that line.
- If a flight test mission is delayed for any reason, the relevant rehearsal needs to be repeated when the previous rehearsal has become “stale”. Experience has shown that this occurs after about 1 to 2 weeks or when key personnel supporting that mission have changed (pilot, test conductor, or discipline leads).
- Flight test support simulators can also be used to supplement general pilot qualification and currency training requirements. This has been accomplished effectively when the local simulator has capabilities adequate for that purpose. This can provide a significant benefit by avoiding the necessity for pilots to travel to remote simulation facilities and keeping test pilots nearby and available to support ongoing test activities.

CA5: INCORPORATE FLIGHT TEST AIDS ON TEST AIRCRAFT

LESSON LEARNED: Large test programs involving extensive envelope expansion testing (flying qualities, flight controls, loads, flutter, etc.), have sometimes neglected to incorporate flight test aids in the initial contract. Flight test aids have eventually been incorporated but as an afterthought leading to inefficient implementation.

PROBLEM: The early planning on large EMD programs has focused on the overall project and missed the important design requirement for flight test aids. When the realization of the need became apparent it has been considered out of scope and therefore required additional unplanned funding. The late timing of incorporating these aids into the design has driven less effective implementation compared to designs that would have resulted if incorporated from the beginning.

DISCUSSION: On large EMD programs, by the time the need for flight test aids was widely accepted as a requirement the overall system design had typically already been established. Because the requirement was defined late it was often viewed as "requirements creep" and was more expensive to implement than if included as an integral element of the design. This has led to awkward implementations with less functionality. An example of an awkward and more expensive implementation has been the requirement for a separate hardware control panel in the cockpit as opposed to integration with existing multi-function displays. The functionality of the tool has been compromised by having to do work arounds to implement the aid outside the scope of the fundamental design. The concept for flight test aids is not new but can be traced to the 1970s. Some related recommendations are included in TIH 91-01 (reference 3).

APPROPRIATE ACTION: The initial requirements for large EMD test programs need to include the fundamental identification of flight test aids so the initial design concepts incorporate these aids. Prior to establishment of the initial contractual requirements, the government program office and flight test community should collaborate to define the top level requirements for a given project. For example, the top level requirements would include the need for the broad categories of types of flight test aids, e.g., flutter excitation system, off nominal control law gain changes, automatic maneuver generation, automatic load factor and rate limiters, etc. Subsequently the government and contractor team should collaborate to refine the specific details of the functionality of the flight test aids.

In situations where it is too late to influence the contract, the flight test community needs to initiate discussions as early as practical to advocate the best plan that can be obtained given the contractual omission. This may be difficult when most other people will be focused on much broader design concepts and programmatic issues such as establishing fundamental programmatic infrastructure. However, this issue cannot be allowed to languish without attention until it is too late to apply a reasonably practical and effective implementation. Flight test engineers will need management help to get this topic established as a "long lead" agenda item at ITT meetings or other appropriate forums. Meanwhile, the flight test discipline engineers should seek allies at similar levels within the program office and contractor design and test organizations. Even when key management in those organizations may not understand the need for flight test aids, it is much more likely that the need will be recognized at the working level. When there is a consolidated position across the entire project discipline level to the effect that, "Yes we need this", it is much easier to determine a plan to identify where the extra funding will be obtained. That extra funding will also be easier to obtain when the need is clearly framed in terms of the projected flight test savings. Those projected savings can be defined in terms of the number of test point repeats that can be avoided, enabling flight test to continue without the need to wait for a software update, and avoiding the inspections and downtime that would come from overloads that would be more common during manually executed maneuvers. The high costs of flight testing should make those savings very obvious to those that may not support the concept initially.

SUPPLEMENTAL DISCUSSION: Some envelope expansion testing requires unique test techniques and maneuvers that can be very difficult to achieve manually, and may lend themselves to more easily exceeding interim limits. The consequences of attempting these tests manually, without flight test aids, are numerous repeats and potential temporary grounding of the aircraft for inspections caused by accidently exceeding interim limits, and delays in flight control development waiting on OFP updates. Fly-by-wire aircraft provide a great opportunity to implement flight test aids that significantly reduce the difficulty of accomplishing these tests, and expand opportunities to refine gains, etc., thereby enhancing test efficiency and safety.

More Specific Examples of Flight Test Aids:

Many test points in envelope expansion testing (e.g., loads, flying qualities) requires hitting a detailed combination of altitude, airspeed, Mach number and g to execute the test point, without exceeding limits in the build-up. A "dial-a-g" capability as a test aid built into the test aircraft flight control system could help the pilot limit the max g to the target g of the test point, limiting one variable of concern, and precluding possible over-g. Other envelope expansion testing (e.g., flutter, flight controls) requires excitation of the system at a certain frequency and magnitude (amplitude) at specific flight conditions to determine system stability, gain and phase margins, and damping. A flutter and flight control excitation system built into the test aircraft flight control system would enable the pilot to select discrete frequencies and amplitudes of excitation, thereby providing the needed inputs at the targeted parameters. Manual inputs at the required frequency and/or amplitude can be quite difficult to achieve. If flown manually, these tests points would most likely be repeated numerous times to achieve the needed parameters of the test point, within the specified tolerances. The result would be more flight test repeats than planned and extended test schedules. The costs for the additional flight testing and extended schedules would far exceed the cost of implementing the flight test aids.

Another type of flight test aid that supports flight control system development testing has been called terms such as "dial-a-gain". When the flight control system is initially designed and developed, simulation is used to refine gains for both the control laws and air data. Although very useful, history has shown repeatedly that the simulations aren't 100 percent reliable for refining gains. The aerodynamics were modeled based on wind tunnel data, the servo-actuators were modeled based on simulation and design specifications, etc. Flight test frequently exposes differences that are present in the real world that weren't predicted or recognized in the design process. Having the ability in flight to adjust certain gains to refine the flight control system design provides a tool that significantly reduces development and flight test risk. In addition, this type of flight test aid provides a limited capability to respond to unpredicted aerodynamics or air data without having to wait for an OFP update, thereby enabling continuation of testing. This provides a significant cost savings compared to the cost of implementing a new gain in the OFP update, only to find that there are still residual issues that require still another OFP update.

Some flight test aid capabilities have required that the production control laws (typically hosted in the flight control computer) be bypassed with the automated command signals feeding the actuators more directly. This has been particularly true for flutter excitation systems that need to move the control surfaces at higher frequencies than would be allowed by the normal filtering, gain scheduling, and limiting provided by the production control laws. Although this type of capability has been very important and may have been essential on some projects, bypassing the production control laws also bypasses the "built-in" safety features that the control laws provide. This just means that the test team needs to be far more cognizant of the implications and establish procedures to ensure that safety constraints are not violated. For example, the amplitude and frequency of automated inputs that bypass control laws could easily exceed structural loads or fatigue life limitations at a given flight condition. A preprogrammed set of control surface amplitudes may be well within loads limitations at a low airspeed flight condition but the same amplitude could be well beyond those limits at a higher airspeed. Those conditions need to be explicitly evaluated and cleared by the responsible structural loads design team to establish ground rules for how and where a given set of preprogrammed inputs can be utilized.

In addition, the combined amplitude and frequency of a given preprogrammed input sequence could be far different than the normal duty cycle of the actuators. The planned preprogrammed input sequences need to be discussed with the actuator vendors very early to find out if there are potential problems. However, the requirements and contracts with the actuator vendors also tend to be established very early so new assessments may be considered out of scope. Ideally, the contract between the prime and the actuator vendor will include support for assessment of flight test aids and their unique conditions. The government probably cannot legislate contractual clauses to that effect, but may be able to recommend consideration to avoid future issues.

A very important type of flight test aid provides the capability to efficiently and effectively control the center of gravity (CG) for flight test purposes. Most modern aircraft have a "built-in" fuel burn schedule that automatically follows a pre-designated sequence for using fuel from specific tanks in a consistent order to maintain a "nominal" CG throughout the weight range. That type of fuel burn system typically works fine for a production aircraft, but does not support the necessary range for a flight test aircraft. Flight test aircraft need to quickly establish any CG value throughout the permissible range as designated in the flight manual. Flight testing needs to confirm the stability and control and handling characteristics at the most forward CG, the most aft CG, and also up to an operationally achievable asymmetric condition. Those extremes are often unattainable using the production fuel burn schedule, but can be achieved in operational applications through other means such as changing internal or external store loadings, failure conditions such as with trapped fuel in a particular fuel tank, failed fuel transfer pumps, alternate cargo loadings, etc. It is usually impractical to intentionally use those methods for establishing an extreme CG for flight test purposes. Therefore a non-production method must be provided (at least on some of the test aircraft) that enables all of the necessary flight test fuel states to be achievable without taking excessive flight test time and without using excessively complicated procedures.

Flight test projects have developed the capability for flight test control of fuel state using a wide variety of methods. Not all of those have been called "flight test aids". Some projects have used a flight test unique fuel control panel to manage CG (different than the nominal burn curve). Other projects have interfaced standard multi-function displays to enable direct control of individual production fuel pumps. Some projects have only added a single switch in the cockpit to burn fuel from forward tanks or aft tanks, thereby enabling the target CG to be achieved… eventually. Other projects have added flight test unique fuel pumps to enable more rapid fuel transfer from forward to aft tanks (or vice versa) or to transfer fuel laterally from one side to the other to obtain an asymmetric condition as needed for specific test points.

Regardless of the fuel management method selected for a given project, a prime consideration needs to be the length of time required to move from one fuel state to another. Many projects have made the decision to save tens of thousands of dollars (typically within the rather narrowly focused budget for a given department) by implementing a less capable flight test fuel management system. However those savings were quickly eroded during flight testing because it could take a half hour or longer to obtain the target fuel state. Given that flight test costs could be as high as hundreds of thousands of dollars per hour, the perceived savings from a limited point of view resulted in a net negative for the overall project. The bottom line is that the flight test costs need to be part of any tradeoff analysis when making decisions regarding a flight test fuel management system. That system needs to be implemented in a way that avoids excessive complexity but enables all required flight test fuel states (including longitudinal CG and lateral asymmetry) to be achieved as expeditiously as practical.

Another category of a flight test aid that should be given consideration is a special indicator on a Heads-Up-Display or other device. One very simple example that has been used with excellent results is a display to show the pilot which direction to move the throttle to obtain a trim point at the current flight conditions (i.e., without accelerating or decelerating). That simple display can save a lot of flight time that would otherwise be used when the pilot waits to notice a small mistrim, adjusts the throttle, and then repeats the process multiple times until an adequate trim point is obtained.

Flight test aids of various types have been implemented on many previous programs. Not every type of flight test aid designed into a given system was used, but the ones that were used typically provided major cost and schedule savings that would have been unavoidable without those predefined capabilities. Some of the flight test aids are designed into the system in anticipation of typical problems from inaccurate predictions. If the problems are not encountered, some of the aids implemented as a contingency may not need to be used. That result should be viewed as a happy occurrence, not as an unused expenditure.

Another important consideration for flight test aids is providing the pilot with the ability to quickly disable the flight test unique configuration and return to a nominal configuration. For example, the pilot should be provided the means to quickly disengage excitation of the structure, or flight controls. If the excitation is producing an unexpected result, the pilot needs the means to quickly disengage the excitation. Another example is providing the pilot with the means to immediately disengage a revised gain in the flight controls if an unexpected response results from the engagement of the adjusted gain. Many test aircraft have employed the paddle switch on the control stick as a means for the pilot to quickly disengage the test aid. The specific disengagement method selected is not critical, but the need to disengage quickly normally leads to the choice of a Hands on Throttle and Stick (HOTAS) mechanism.

There are exceptions where it would be unwise to allow disengagement of some specific flight test aids. For example, if a gain reduction is necessary to allow continued envelope expansion using a flight test aid (instead of waiting for a more formal OFP update), it would be unwise to allow disengagement of that gain at the same time as disengagement of a flutter excitation sequence. Disengagement of the flutter excitation could be warranted by an unexpected response, but disengagement of the gain reduction and returning to the original (excessively high) gain could cause a worse situation. The whole point of this concept is for the test team and designers to carefully consider which mechanization makes the most sense for their particular situation.

Another concept that has often been used is to provide the pilot with a display that provides sufficient insight into what can be expected when a particular flight test aid is initiated. For example, the pilot may not need to know the detailed characteristics for a frequency sweep, but that pilot needs to understand if that excitation will be long or quick, which surface(s) or axis will be excited, and have some idea of the amplitude. Some projects have only provided the pilot with a single number on a display to represent a complex excitation sequence, but that single number could also represent a wide variety of flight test aid types. If it is truly necessary to limit the display capability for flight test aids to such rudimentary indications, it will be necessary to compensate in some way. A common method has been to put extra information into the test cards, but that adds complexity and may actually cause confusion. In general it will be better to provide the pilot with adequate info via simple but informative cockpit displays.

CA6: PLAN FOR REALISTIC NUMBER OF SOFTWARE REVISIONS

LESSON LEARNED: Flight test programs involving software intensive systems have often drastically underestimated the number of software revisions required to complete the program. A much larger number of software revisions than typically planned are required in order to address normal developmental issues. This leads to very disruptive software limitations, interim operating limits, and highly inefficient flight test work-arounds that stay in place for many months until a satisfactory solution is provided in a software revision (which can often take multiple iterations). Those flight test delays result in longer schedules and higher costs than originally estimated. Although this Lesson Learned is focused on software revisions, the impacts can be exacerbated when flight test configurations can ONLY be provided as complex combinations of both hardware and software.

PROBLEM: On large EMD-level programs involving the development of the entire weapon system, the number of software revisions that are actually required to complete the development has ranged from 2 to 10 or more times the number originally estimated. The result is a substantial increase in the amount of laboratory (lab) and flight testing required. Most programs don't have the additional time and money available because the additional software revisions weren't planned and accounted for. On large EMD programs involving high tech systems, the technology being implemented is usually much less mature than originally estimated, and the degree of difficulty to mature the technology is even more drastically underestimated. Sub-systems that functioned well in a stand-alone environment do not always blend smoothly into a complex integrated system. This problem is not limited to envelope expansion. It applies to all software intensive systems.

DISCUSSION: Classic quotes:

"If we run into minor problems we'll just use workarounds or do something else until the next major software update."

"If we run into a show-stopper problem we can pump out an emergency update in less than 2 weeks."

Both of those phrases have been heard in various forms on multiple large scale EMD-type projects. Both have tended to be an indication of unwillingness to build-in reasonable contingencies aimed at precognitive plans for solving problems. It has been common for the person expressing those words to sincerely believe them. The first phrase might have had some validity in the hypothetical situation where there were only a few minor problems. However, in the much more typical real-world situation when there were dozens of major and minor problems across multiple systems, the ability to rely on workarounds or alternate paths has become completely impractical.

The good news about the second phrase regarding show-stopper problems is that previous projects have proven it to be true… over-and-over again. Each time a show-stopper problem was discovered, it has been truly impressive how quickly a fix was designed and implemented. The bad news was that each of those emergency updates was unplanned and could only be delivered at the cost of a similar delay to the next planned major update (also delaying urgently needed new capabilities).

This Lesson Learned is focused on Air Vehicle software (flight controls, propulsion, subsystems, etc.) but the lesson also applies to the broader system under test (e.g., avionics, offensive systems, defensive systems, sensor systems, etc.). The difference between the number of planned and actual software revisions required in a flight test program is often related to the complexity of the intended capability, the maturity of the technology involved, and the degree of difficulty to advance the maturity of the technology. Developments involving existing systems/capabilities that are enhancements of the capabilities without significant advancements in technology usually require fewer software revisions and the original estimates tend to be more accurate. Developments of large complex, integrated systems involving advanced technologies usually significantly underestimate the number of software revisions required. The maturity of the technology is overestimated, the ability of sub-systems to function as intended in a highly integrated system is overestimated, and the degree of difficulty to advance the maturity of the technology is very much underestimated.

A serious cost and schedule impact that has often been overlooked when implementing optimistic software release schedules is the cycle time between discovery of a problem in flight test and the installation of an OFP with an intended fix for that problem onto a test aircraft. That cycle time has often been 6 to 9 months or longer. The reason it tends to take so long is because there are so many steps in the process to design and implement a fix. Those steps typically involve: identification of a problem in flight test, confirmation that it was actually a problem (versus a misunderstanding), identification of the root cause, design of a solution, identification of a target OFP for that fix, coding of the intended solution, verification of the fix in lab testing, certification of the overall OFP (often implementing many other fixes), re-test in flight. If each of those steps can be done in a single day, it is hypothetically possible for the fix to arrive in a week, but it is far more common for each step to take days or months. If that fix does not work out as expected, the cycle time resets to another 6 to 9 months until another OFP arrives at the test site. In addition, if the problem was discovered within a few months of the planned release for the next major update, the design fix would probably miss the freeze date for that update and would automatically be relegated to the following planned major update.

Every EMD project has acknowledged the need and capacity for an occasional "emergency" flight test OFP update capability. These emergency updates are advertised to provide a fix that can be developed, verified and validated through lab testing, and delivered to a test aircraft within just a few weeks. However, those cases tend to be reserved for true schedule emergencies in which flight testing cannot continue at all unless that fix is delivered. The concept of emergency updates does not support the normal development process that occurs on every major EMD project. Instead of regularly providing emergency updates, it is far more normal for interim operating limits and inefficient work-arounds to be implemented that allow flight testing to continue, albeit on alternative paths that may not be the critical path. An emergency OFP update is typically judged to be impractical because it would delay efforts on OFPs that are already in the pipeline and needed to deliver important new capabilities. When the next planned OFP is already too far into its development cycle, the growing list of flight test fixes is delayed for insertion into the second planned OFP, or even the third planned OFP in the pipeline. Continued flight test progress often depends on heroic efforts on the part of the design organizations and the teams actually accomplishing the Verification and Validation (V&V) efforts in the labs. Nevertheless, the end result has been many months between identification of a problem and the first opportunity to evaluate a fix in flight test. See also: "CA8: Tracking Flight Test Progress" and "FF5: Providing Focus on Fixing Flight Test Deficiencies" for related concepts.

In an ideal world, every EMD project would have the capacity for sufficient OFP pipelines so that flight test fixes could be introduced quicker without delaying planned OFPs. However, the additional lab and people capacity will always be perceived as unnecessarily costly given that all systems are expected to function very well without the need for multiple updates. Typically the project is assumed to have far more maturity or to be far better managed than any of the historical examples. Every project makes the judgment they are in that category until the facts indicate otherwise, at which time it is too late to take any practical actions other than to hire more people and surge into a multi-shift OFP development pipeline. Even if two separate OFP pipelines are in operation, that approach would likely lead to configuration control issues where the flight test fixes are in a different pipeline than the next planned OFP, creating confusion on the actual capabilities of delivered OFPs, along with uncertainty on which OFPs are subjected to which interim operating limits.

Recent legacy programs on new fighters and bombers provide a sense of the number of revisions required versus planned for a very complex high tech weapon system. The typical pre-EMD software plan on recent EMD projects has been an initial software release to "get things started," a mid-EMD release to "add new capabilities and fix any major bugs with the initial version," and a final release to "complete the capability suite and apply final tweaks to the previous version". The actual number of software revisions on those legacy projects was many dozens (with about half being major revisions), not just the three capability-oriented OFPs that were initially envisioned. That same trend has not been limited to large EMD-type projects, but has also been observed on relatively moderate-sized development projects on legacy aircraft and Foreign Military Sales efforts.

The impacts of all these unplanned software releases are not limited to just air vehicle envelope expansion flight testing. Avionics, Electronic Warfare/Low Observable, or weapons tests that require certain portions of the flight envelope may not be accomplished on schedule either, if software revisions are required to clear those portions of the envelope. This multiplies the negative impacts of the original issue, since discoveries of anomalies with those other systems are also delayed, not to mention subsequent corrections and software revisions to address those anomalies. Unplanned software releases are prevalent on all software intensive systems which also highlights how critical it is to have a reasonable system wide software revision plan up front.

APPROPRIATE ACTION: There is no easy solution when planning for a realistic number of software revisions. No project will have 20/20 foresight when predicting the actual need. However, instead of starting an EMD project with a highly optimistic software release schedule, it would be better to plan for more realistic, regular software releases throughout the course of EMD. A realistic software release schedule would include at least 3 to 4 updates per year. Every EMD project accomplished over the last three decades has far exceeded that regular software update pace when the flight test realities were encountered. Any software release plan for a new EMD project with fewer updates than the proposed realistic pace should be viewed with great suspicion and should be aggressively challenged by the flight test organization.

Even if the proposed realistic pace turns out to be pessimistic and unnecessary, it will be much cheaper and less disruptive to cancel or consolidate a few software releases than it would be to insert unplanned releases. Once a more realistic software release schedule is developed that also includes more realistic estimates of software release cycle times, it should be baselined to allow all stakeholders to better understand when the required flight test OFP delivery rate begins to exceed the original baseline. Conversely, if the required flight test OFP delivery rate is less than the planned baseline, that would be an excellent metric to make a case to senior acquisition officials that the project is indeed ahead of pace. Another important metric would be cycle time for fixes to problems discovered during flight test. If that cycle time can be regularly provided within 2 to 3 months from identification of each problem, it would provide a really good indication that the project has improved on historical norms. The tradeoff between those two aspects should be at the heart of EMD OFP development efforts. Fewer OFPs can be delivered each year, but at the probable expense of cycle time on flight test fixes. The "art" will be in defining an optimal balance for a given project.

SUPPLEMENTAL DISCUSSION: This Lesson Learned has the most significant impact on non-flight test programmatic aspects (overall cost and schedule). Some of the additional unplanned impacts have included: retention of a larger design team, retention of a larger software test team (or additional hiring to expand capability for additional work-shifts), reconfiguration of scarce lab resources along with the associated disruption to planned software release schedules, development and Verification and Validation retest in the System Integration Lab (SIL); in-flight regression tests to ensure the changes have not introduced unintended consequences; and flight test repeats of the test points that uncovered the problem to prove that the problem has been fixed. All of these factors extend the schedule and increase costs significantly beyond what was estimated. Therefore this issue would be most properly addressed between the Program Office and contractor with oversight from higher levels such as DASD (DT&E). However, there are still important flight test implications that should be of concern to any test organization.

CA7: PLAN FOR SHIFTING WORK SPLIT WITHIN TEST TEAM

LESSON LEARNED: The test team work split on large EMD development programs is typically planned for a high level of contractor support throughout the program. However, most programs have tended to be staffed mostly by the contractor in the early portions and mostly by the government in the later portions. This has created significant disruption to other projects when the government is called upon to staff at higher levels than planned.

PROBLEM: Although the contractors fully intend to support the EMD effort at a high level throughout the program, they experience difficulty attracting the right types of people to relocate to the test site for the duration. In the early stages of the EMD program the most experienced contractor people are on site. As the program evolves, these individuals are reassigned to other projects. This is the phase of the program where the government is called upon to staff at higher levels than planned. This is also the phase requiring the most expertise because testing is reaching the envelope extremes. The unplanned government staffing comes from other programs that are consequently impacted by the loss of personnel.

DISCUSSION: Large EMD type development flight test programs have been much more successful and efficient when there is a well-integrated team of dedicated contractors and government testers on site working the test program and all its related issues. The contractor brings expertise in the detailed system knowledge, lab test experience, sometimes prototype flight test experience, and the insights of making things happen within the company. The government testers bring the expertise of flight test, broader flight test experience, a perspective from the operator, and the breadth and depth of lessons learned from previous similar projects. This well integrated blend of expertise results in more efficient testing and more timely resolution of problems. This is why planning for a sustained high level of contractor support is a good idea. However, a variety of factors have tended to prevent this from happening.

One of the primary factors that has undermined the plan of maintaining a high level of contractor support is the inherent reluctance of people to relocate to the test site for several years. Attractive incentives that persisted for the duration of the relocations may have helped offset that inherent reluctance during the earlier portions of an EMD project. However those incentives have tended to fade over the course of EMD. In the later portion of some programs it has been surprisingly common for companies to expect their people to relocate at their own expense and without any kind of Per Diem once at the test site. Unsurprisingly, volunteers for those situations have been hard to find.

Another factor has been the reassignment of the most experienced contractor personnel. Once a routine flight test tempo was achieved, the most experienced contractor personnel tended to be reassigned to other priorities within the company. In addition, some contractor engineers have highlighted the fact that high performance contractor personnel tend to avoid spending much time deployed away from their home organization because it would be damaging to their careers and limit promotion opportunities.

It has also been common for companies to rely on hiring of longer term personnel locally at the test site location. That method has tended to provide intermittent success based upon the overall state of local hiring by the government and other contractors. Depending on the timing, the most experienced and qualified personnel had often already been hired by those other organizations. Also, the locally-hired people tend not to have the design experience on the given system that their counterparts from the plant itself will have, lessening the positive impact that contractor engineers typically bring to the effort.

One consequence of these contractor staffing situations has been that the contractor resorted to remote support or frequent rotation of personnel (typically about 2 weeks). This has tended to dilute contractor contributions to the ongoing flight test effort. Although those rotating personnel were often very well-qualified, they were not dedicated to flight test support. Instead of helping to "turn the flight test crank" by working on the many tasks related to preparing for upcoming test missions, those individuals often felt their priority was to work on "home office" tasking. If the individuals assigned to rotational duty weren't familiar with the flight test environment, or it had been a long time since their last exposure, they might just be coming up to speed when it became time for them to return to their home office. This also tended to dilute their commitment to flight test related tasking.

None of these factors are within the prerogative of the government since they are driven by contractor internal policies. However, recognition of this historical trend may allow the government to be better positioned by planning for that eventuality. For example, another consequence of this contractor staffing shift is the upheaval it has caused in other local programs when the government has had to compensate for the reduced contractor staffing by pulling highly experienced people from those other programs.

APPROPRIATE ACTION: Large EMD programs need to recognize the staffing realities that typically play out in those programs and plan accordingly. The reality has been that the contractor staffing was sufficiently high in the early phase of EMD and became less than planned later in the program with a corresponding unplanned (and often dramatic) increase in government staffing. The likelihood of this occurring needs to be recognized by government staffing organizations and possible mitigations planned in anticipation by ensuring the right experience levels and skill sets are available. That can be tricky when all of the "official" programmatic schedules are based upon unrealistic staffing assumptions.

Prior to contract award it would be appropriate to advocate for a stable test team staffing plan for both contractor and government personnel. It may also be possible to advocate for a staffing plan that fits a more realistic historical trend. As a starting point for discussions (depending on the type of project), it can be quite reasonable for a joint staffing plan to be somewhat biased towards contractor support initially. Contractors will always bring the most system knowledge to the team during the critical early time period. A reasonable staffing plan would then shift fairly rapidly to a normal, healthy test team with something approximating a 50-50 split as the government portion of the team becomes more familiar with the system under test. In order to be representative of historical trends, the plan should be more heavily staffed by government personnel in the last half of EMD. Naturally, the exact ratio has been variable but it has not been uncommon for the government staffing to end up being 70 percent or higher.

Even after contract award, it is still reasonable to acknowledge and plan for heavy contractor staffing initially. In particular, staffing for the first flight mission control room team has been a point of contention on several projects. Contractors have often felt that the first flight control room team needs to be staffed by their people exclusively. The government has often aggressively advocated for a much stronger presence on that team. As a general principle, the most qualified individuals should staff the first flight control room, regardless of their organization. It is reasonable to expect that at first, those people will mostly (but not always) be from the contractor since they will probably have the most system knowledge. The primary concern should be to prevent that team from being too heavily dominated by designers with little or no flight test experience. If the first flight team has a reasonable balance between knowledgeable system designers and experienced flight testers, it shouldn't matter if most of them are wearing company jackets. Many government engineers and test conductors need to have more patience and restrain their enthusiasm by allowing the company reps to bring their own "baby" into the world. There will be plenty of future opportunities. If the contractor acknowledges the need for a more experienced government person to support a given position on that team, the government should be ready and willing to provide that support. A really good way for government engineers to earn more respect from the contractor team is to spend more time at the contractor location long before first flight. There are usually many opportunities to participate in design discussions, lab and simulator testing, test plan working groups, etc. Naturally, this would need to be done in a way that is not perceived as a contractual "burden" to the contractor, and not regarded as tactic for identifying flaws that can then be reported up the USAF chain, but as a team building process with opportunities for genuine contributions. RFPs often provide a contractual basis for this type of participation but government flight test engineers have not always taken sufficient advantage of these opportunities. This has sometimes occurred as the reverse of the situation where contractor star performers want to minimize time at the flight test location. Government staffing estimates need to account for people spending significant time at the contractor location.

Even after first flight, it is still reasonable for the mission control room team to be heavily contractor-oriented. However, it is also reasonable for the government team members to begin participating within each technical discipline or control room position. In disciplines with multiple control room positions, over half can still be staffed by contractor reps. The goals of the government people should be to improve system-specific knowledge and prepare for more challenging future missions. Early involvement of government people in lab and simulation testing can help get them higher on the learning curve sooner. For control room positions that only require a single person (such as Test Conductor or Test Director), a rotation can be implemented, ideally with oversight from a more experienced person in that role.

Based on past experience, the absolute most effective test teams have been completely interchangeable between contractor and government personnel at any particular station in the mission control room. When the first few missions have been successful, the goal should be to achieve that interchangeable team as soon as practical. Once into an effective flight test rhythm, the assigned test teams have been selected based on an individual's skills and availability, not on organizational affiliation. If a given individual was needed to prepare the path to upcoming missions, someone from the other organization would fill the control room role. That is a very GOOD plan that should be advocated throughout the planning cycle.

CA8: TRACKING FLIGHT TEST PROGRESS

LESSON LEARNED: Tracking flight test hours flown and test points accomplished as the primary indications of flight test program progress does not provide meaningful measures of actual progress or of work remaining on the test program for adequate program control.

PROBLEM: Flight test programs typically use flight test hours flown and test points accomplished as measures of progress of the program. Neither of these measures provides sufficient insight and understanding of the actual work accomplished vs work that was planned to be accomplished, nor do they provide accurate insight into the amount of work remaining to be accomplished. The result is either false indications that the program is not on track, or false indications that the program is on track, when in fact a bow wave of work is building that will overwhelm the schedule later in the program. These measures assume effort means progress and lack the measurement and control needed to effectively manage the program.

DISCUSSION: Classic quote:

"Don't confuse activity with achievement." – John Wooden

That wisdom from Coach John Wooden superbly captures the fundamental problem with reporting flight test progress in terms of test points attempted and overall flight hours flown. Those metrics may provide an indication of how active the flight test community has been, but do not provide a useful indication of what has actually been accomplished. Not all test points are equal. Some test points require just a few seconds to accomplish, while others may take an entire test flight for one test point. Lumping all these test points together and tracking percent complete provides a very misleading picture of work accomplished and work remaining. The overall hours flown and the hours remaining are also misleading for the same reason. In addition, without addressing the achievements to date in a capabilities-based context, the question remains: what has actually been accomplished?

An approach that addresses the overall performance problem is normalizing the test points and prioritizing and tracking them in work packages built around capabilities-based objectives and milestones, similar to a work breakdown structure (WBS). The test structure is developed to the work package level, and the interrelationships between test objectives, hardware and software requirements within and between the work packages are defined and developed. This allows scheduling the work and establishing priorities based on required test events, provides a projection of when program milestones will be complete, and visibility into the latest period the aircraft must be configured to support required testing. This approach has been used very successfully. It measured the work scheduled and the work performed which improved the process of flight test planning, measurement, and control. The test points were normalized based on an estimate of the amount of time required to accomplish that particular test point (test point-hours). This time included time to get on conditions, actual execution, and recovery. This provided a direct cost of gathering the data. The time required to accomplish each individual test point was determined by the engineering disciplines.

It was important that the determination of the time was "vetted" by test engineers who had more recent experience in test execution. Designers tend to provide optimistic estimates that could be misleading. The flight hour requirement for the program was determined by the sum of all the test hours determined for each test point divided by a flight hour effectiveness factor. This enabled accounting for overhead aspects of flight test such as takeoff, landing, transitioning to test airspace, etc., in addition to estimating for a number of repeat test points due to regression testing, bad data, etc. Having the test points grouped in work packages built around capabilities-based objectives and milestones provided a ready measure of progress against specific capabilities and milestones and the requirements to achieve those capabilities and milestones. See Appendix B for more details of the system used successfully on a past program.

The accounting process must be disciplined and the flight hour effectiveness factor must be realistic for the technique to work. When a test point has been flown, the test point status is "pending". The engineers look at the data to assess if the data is good and the test point is complete (does not need to be repeated). If complete, it is noted as so in the tracking system and credit is taken for work accomplished and completed. If it is not complete, it is noted as so, and work is shown as having been accomplished, but not completed.

In addition, if there are system changes necessitating repeat or regression testing, the new work is accounted for by noting the test points (test point-hours) that must be repeated, which adds to remaining work. That added work can be due to government requirements changes, contractor initiated design changes, or due to implementation of fixes to deficiencies discovered during prior flight testing (see FF5). The entire category of added work will need special attention to properly document since those additions tend to happen at a time when the most senior project leadership (both government and contractor) very much want to see the remaining work going down, not increasing. Everyone wants to take credit for implementing fixes to problems, but no one wants to get blamed for expanding the flight test scope. In order to have any chance of realistically managing a very complex flight test effort, any tool for tracking flight test progress needs to properly capture a running assessment of the amount of actual progress, the amount of effort expended, and the amount of work to go.

Some useful points of reference from Lesson Learned "CA3: Expected Test Efficiency for Test Program Planning."

- The flight hour effectiveness factor (i.e., the percent of flight time actually spent executing test points) typically runs in the 60 percent range (0.60).
- The percentage of test points that need to be repeated over the course of a program has exceeded 50 percent on several projects. That high number of repeats has often occurred because multiple design fixes were necessary before deficiencies were adequately addressed, requiring the same test point to be repeated many times.

APPROPRIATE ACTION: Large flight test programs should employ a test progress tracking system that enables a more accurate tracking of the actual progress and work remaining in a flight test program. An example of a system that was employed with great success is described above and in Appendix B. It provided a more credible and meaningful report of flight test status for the Program Office, Service Department, and DOD Acquisition Manager. Each new project will need to adapt the methodology to be more compatible with program philosophies, but the basic principles should be applicable to almost any flight test program.

A normalization technique that provides a realistic estimate of the time on condition required to accomplish any given test point is at the heart of the methods described in Appendix B. That technique provides a relative time cost for each test point that helps avoid the misperception that each of those flight test "beans" is the same. Therefore the fundamental "bean" being tracked is in units of time, not units of test points. The resulting metrics provide a much more realistic understanding of the work performed and work remaining, regardless of the particular format for those metrics. However, those metrics are still mainly an indication of flight test activity and do not directly provide a sense of accomplishment of significance to the overall project. Just because a bunch of flight testing occurred doesn't mean the system works as intended.

A real measure of accomplishment is best achieved using capabilities-based work packages. Each work package is focused on an incremental achievement of importance to that development project. Examples of the work packages should be descriptive and understandable and may include things like: fully clearing the middle of the flight envelope, successful completion of all testing at 80 percent Loads, completing testing at an envelope extreme in some dimension (max Mach number, max dynamic pressure, max load factor, max gross weight, etc.), successful separation of a key weapon (missile, bomb, or bullet), completion of all aerial refueling flight testing for a particular tanker aircraft, successful demonstration of a single engine air-start, completion of all engine air-start testing, successful demonstration of a Key Performance Parameter, etc. When each of those work packages has an estimate for the total time on condition of all associated test points, the relative cost for each work package is much more obvious. The work packages of the highest importance are most effective when tied directly to fundamental capabilities as described in the System Requirements Document or other user-oriented documents.

CA9: UPDATE SYSTEM MODELS WITH FLIGHT TEST DATA

LESSON LEARNED: On almost every new aircraft program, flight test data has not been adequately used to provide comprehensive updates to the system models as needed to support ongoing development and sustainment. This misses the most significant aspect of the value of the flight test. Updated system models would provide more accurate data for the flight manual, lower risk and cost of DT and OT flight testing by allowing more reliable predictions for interpolation/extrapolation within the flight envelope, lower risk and cost in the resolution of system development problems, and lower risk and cost in subsequent developments of that weapon system, thereby providing significant leverage from the money invested in flight test and model updating.

PROBLEM: Flight test data is rarely used to do comprehensive updates to the system models that are employed for developing fixes to existing problems, generating flight manual data for aircrew use, providing aircrew training simulations, and subsequent developments of the weapon system capabilities and life cycle extensions. This leads to using models with sporadic updates at best which produce less reliability in their predictions than would be available if comprehensive updates were done with flight test data. This results in higher development risk, with corresponding higher costs.

DISCUSSION: Classic quotes:

"Of course we'll update the model, but only if the differences are big."

"This sounds like a science project to me."

Both of those phrases have been heard in various forms on multiple large scale EMD-type projects. Both should be interpreted as danger signs indicating future problems. These phrases indicate that the perceptions of the importance for model updates varies at all levels. The first phrase tends to come from middle management and the initial intentions may be good, but budget cuts and priority issues seriously degrade the final product actually delivered. The second phrase tends to come from more senior management and reflects a general lack of technical depth. Both phrases indicate that it may be very difficult to find support for a reasonably useful model update process from within a project. Fortunately, recent acquisition initiatives have shown that senior USAF leaders have recognized the historical pattern and have attempted to take steps to improve the overall situation. These initiatives are called "Digital Thread" and "Own the Technical Baseline" and are discussed in the context of Lesson Learned "CA11: Government Access to Fundamental Aircraft Models." These initiatives address the importance of having access to updated aircraft models that reflect flight test results.

System models are used extensively to support system design and predict system performance. These models include aerodynamics for performance, flying qualities, and control system interaction predictions; propulsion for thrust and engine stability predictions; structures for loads and flutter predictions; flow field effects for weapon separation; aero acoustics for weapons bay environment characteristics predictions; radar cross section predictions for low observables; braking system for performance predictions; environmental control system for performance predictions; sensors, offensive avionics, defensive avionics, and navigation system for mission systems predictions and effectiveness studies; etc. These models are generated from sources that provide a theoretical representation of that aspect of the system, which produces predictions with various levels of credibility. Hence the need for flight test. When flight test is conducted, data is produced that provides the real characteristics of the system.

Contracts with weapon system developers often do not include the requirement to update the system models with the flight test data until the very end of EMD, if at all. The result is that the original (or partially updated) system models continue to be used for resolving system development issues, providing flight test predictions, providing system performance data in the flight manuals for aircrew use, and for subsequent system developments. The money spent for flight test is not leveraged by updating the models throughout EMD so that subsequent predictions are more accurate, more reliable, reducing development risk, providing aircrews with more accurate performance data, reducing subsequent flight test scope and risk, and possibly precluding the need for repeat flight tests later in the system life cycle.

Case in point: On a fighter development program, flight testing was conducted to measure the performance (cruise, climb, turn, range, fuel consumption, etc.). The model for the fighter performance was not updated. The flight manual performance predictions were based on the original aerodynamic predictions that had been determined to be inaccurate from the flight test. During Desert Storm, aircrews were experiencing significant issues with mission planning that required very tight fuel consumption specifics in order to meet mission requirements and refueling rendezvous. Serious issues were encountered with low fuel states that hampered and confounded mission execution. The Program Office determined that the contractor did not have the means to correct the data and requested a quick reaction flight test program to determine the performance characteristics to update the flight manual. The Air Force Flight Test Center (now 412th Test Wing) had generated a report on the measured performance from the original flight test program, and provided it to the program office again. The performance documented in the report matched what aircrews were experiencing. The flight manual performance charts were subsequently updated with data produced by the model updated with the flight test data from the original test program, negating the need for additional flight tests.

One of the most consistent problems that has resulted when aircraft models were not updated to reflect flight test results has been the occurrence of residual oscillations (See “FF8: Residual Oscillations”). This has often occurred because aerodynamic models were judged to be “close enough” based on flight test results in one configuration. Even though differences from predictions were visible in control surface effectiveness or another key aerodynamic term, those differences were not significant enough to cause the aircraft to miss a requirement (such as meeting Level 1 handling qualities criteria) in that configuration. Therefore the decision was made not to update the aerodynamic model. However, when a different configuration was flight tested those same observable differences in aerodynamics were sufficient to lead to residual oscillations that delayed the test program by several months before an adequate resolution was available. In hindsight, those oscillations would have been entirely predictable had the models been updated and used to generate a new set of flight test predictions. That same pattern of un-updated models leading to unpredicted problems has been applicable to numerous other technical aspects in addition to residual oscillations.

APPROPRIATE ACTION: Each new project will need to find the right balance between accurate updated models and “science projects”. Contracts for new weapon system development projects should include regular updating of the system models and simulations with flight test data, particularly those that did not correlate with the flight test results. This type of contractual language should also be included in sustainment contracts for system capability updates, as applicable. The updated system models and simulations need to be included in the contract as a recurring deliverable over the course of the test program to ensure the models are updated regularly and are available for ongoing weapon system development. A single set of model updates provided at the end of EMD will not have any useful impact for ongoing flight testing throughout EMD. Planned updates every 12 months throughout EMD may be a workable rate, whereas a high update rate like “every 3 months” seems unnecessarily frequent. An additional clause for “As needed to support ongoing flight test” would recognize the need for model updates required by special circumstances.

Flight testers should be very wary of highly segmented model updates with fuzzy boundaries between regions of applicability. Ideally, each model needs to be provided as a single version that works in all portions of the envelope tested to date, without odd “cut-out” regions where the user needs to turn on or turn off flight test increments. This can be a very difficult goal when limited flight test data is available and it can be tricky to "blend-in" a known flight test impact into a neighboring region (i.e., does that same flight test increment apply or not?).

In addition, it would be beneficial for the contractual language to explicitly identify the expected fidelity for each model or simulation. That fidelity definition in the contractual language would be best described in broad terms related to the expected capability, not detailed technical terms. For example, a broad description of the accuracy expectations may be something like: "The performance model shall predict the fuel flow at a given flight condition within XXXX lbs per hour of the flight test results at the same conditions", or "The piloted simulation shall predict the departure boundary for a given type of maneuver within X deg AOA of the flight test results for that maneuver", or "The piloted simulation shall qualitatively reflect unique handling characteristics encountered during flight testing". The overall goal should be to provide some guidance for the customer's expected accuracy for the various models and simulations instead of relying completely on a contractor's subjective judgment on how close is "close enough". More specific definitions for the accuracy of internal components of models (such as individual terms in an aerodynamic model or engine model) can be deferred to less formal discussions within the discipline-oriented working groups.

If it is already too late to influence contractual language regarding accuracy of models, the same effect might be achievable by defining similar accuracy criteria within discipline working groups. Obviously that approach would not be as formal or as binding as a contractual definition but it would at least provide a precognitive definition of expectations instead of the more typical differences of opinion that have occurred after the fact on many projects.

Another action that should be considered on future projects is reporting on areas in the flight envelope where a given model or simulation is known to be inaccurate compared to flight test results, while also reporting on areas that accurately reflect flight test results. That type of reporting would be best if included as part of existing technical reports, most likely as an appendix. This reporting would not need to reflect every intermediate update, but would only need to occur at the end of a significant flight test phase. The simple act of identifying these aspects and quantifying the impacts may be sufficient to avoid some of the historical examples where models and simulations should have been updated but were not.

Note that this proposal should NOT be interpreted as a simulator certification or validation effort. In the 1980s, one of the assigned tasks for 412 TW Engineering (or the organizational precursor) was to certify that simulations were sufficiently accurate compared to flight test results. These were known as "SIM CERT" projects. Those SIM CERT projects absorbed far too much staffing and distracted from normal flight test activities and were eventually curtailed. Other USAF organizations have inherited the tasking to certify and validate simulations. However, it would still make a lot of sense for flight testers to report on aspects of models and simulations when they have great familiarity with those aspects as a normal part of doing business. In fact, since flight testers will see how well predictions match flight test results on every single test mission, very few people will have a better understanding of the areas where those models do or do not need improvements. These simulation accuracy appendices should not be perceived as comprehensive since the flight testers should only report on aspects they have directly experienced. Nevertheless, it would be highly appropriate to report those results so that the other USAF organizations could benefit from that knowledge instead of needing to rediscover the same aspects.

In extreme cases of known simulation inaccuracies with significant operational impacts, it may be appropriate to document those cases in a USAF Watch Item or Deficiency Report. Local flight test management would need to make that decision based upon programmatic philosophies and the nature of the impact.

CA10: FLIGHT TEST FATIGUE FAILURES DUE TO FLIGHT TEST EXPOSURE BEYOND OPERATIONAL DESIGN USAGE

LESSON LEARNED: Envelope expansion test programs typically do not account for the increased amount of time the aircraft is exposed to the extremes of its envelope to complete a full flight test program (including flutter, loads, flying qualities, propulsion, vibroacoustics, etc.), which has resulted in fatigue failures in flight.

PROBLEM: Envelope expansion testing generally requires a significant amount of time to be spent at the edges of the envelope (close to the design Mach, airspeed, dynamic pressure, and load factor limits) while performing high-speed flutter, loads, flying qualities, vibroacoustics, and propulsion tests. In contrast, the cumulative amount of exposure time to these extremes considered in the design of the aircraft for the operational environment tends to be insignificant compared to the amount of exposure accumulated in flight test.

For example, although it can take a considerable amount of time on condition to clear an envelope with weapon bay doors open, it would be atypical for an operational aircraft to fly for extended periods of time with the weapon bay doors open. As a result, there can be a significant disconnect between the design usage time and the amount of time required to clear an envelope via flight test.

DISCUSSION: Air vehicles are typically designed with a specified operational service life (as an example, 8,000 flight hours or 30 years). Furthermore, the anticipated usage of the aircraft during its lifetime is based on projected mission profiles (training mission utilization, peacetime mission profiles, combat mission profiles, etc.). The design distribution of flight hours by Mach number and altitude are derived from the projected mission profiles. As a result and as a function of the projected mission profiles, most air vehicles are not designed to spend any significant amount of time at the edges of the envelope.

Conversely, flight test air vehicles are required to spend a significant amount of time at the edges of the envelope in order to complete envelope expansion testing. For example, a single flutter test condition typically requires a slow acceleration onto condition, a 60 to 90 second sweep or random flutter excitation followed by a series of burst or dwell flutter excitations (lasting approximately 5 seconds per burst or dwell), and maneuvers (sideslips or wind-up turns). Although those flutter excitations are necessary to obtain flight test data to validate the design, they can introduce power at frequencies not normally encountered during operational usage and can therefore degrade fatigue life faster than expected. Flying qualities and loads test points (consisting of doublets, rolls, sideslips, and wind-up turns) would typically be integrated with the flutter testing to form a leading edge integrated test block (LEITB). The LEITB may take 5 to 10 minutes of dedicated testing (excluding time for test point setup and data review). In addition, because testing at the edge of the envelope may have negative specific excess power (P_S), multiple test runs may be required to complete a single test condition. Once that single test condition has been cleared, the entire process needs to be repeated at multiple high speed conditions.

In addition, equipment and panels located in and around the weapon bays are generally exposed to an increased vibroacoustic response when the weapon bay doors are opened. The cumulative exposure time depends on weapon bay usage. Typical operational usage dictates that the weapon bay doors are only opened for ordnance releases (or simulated ordnance releases), and are quickly cycled closed after the ordnance is released (resulting in minimal or no design usage time at the highest dynamic pressure conditions). In contrast, it is not unusual for a flight test aircraft to spend in excess of 20 minutes at the highest dynamic pressure conditions with the weapon bay doors open in pursuit of test points. The same test aircraft is often required to accomplish a similar test sequence at multiple altitudes at the highest dynamic pressure (as a result of changes to the structure or flight controls), resulting in total exposure time well over an hour. Prolonged exposure during flight testing to other stressful parts of the flight envelope (such as transonic or in high turbulence) can also contribute to surprisingly rapid accumulation of fatigue cycles.

As a result of these disconnects, flight test programs have historically had a history of in-flight fatigue failures. Even if there was a generalized acceptance that such failures could occur, the specific fatigue failures have always been a "surprise" to the flight test team, resulting in significant delays to the flight test program while root cause determination was conducted, fixes and go-forward plans were proposed, fatigue life tracking programs initiated, unexpected event amendments written, and approval granted to continue testing. If the test team is better able to anticipate the potential for these events, they should be able to prevent many of the delays that have occurred in past flight test programs due to unexpected fatigue failures.

It should be noted that air vehicles are typically required to have a factor of two applied on the exposure time derived from the service life (an aircraft with an 8,000 hour operational service life should be designed to withstand 16,000 hours), and an uncertainty factor of 3.5 dB applied to the predicted aeroacoustic sound pressure levels. These margins are in place to account for uncertainties in the air vehicle design and analysis, manufacturing tolerances, and small deviations in the usage environment. It would be erroneous to assume that these margins will account for the significant increase in exposure as a result of flight test.

APPROPRIATE ACTION: Prior to initiation of the flight test program, fatigue critical components should be identified (via analysis and preliminary ground testing), and a plan to measure and track the actual fatigue life usage of those components throughout the flight test program should be developed. In addition, a plan should be developed to manage the fatigue critical components at the flight test location. For example, define any required periodic inspections, have replacement parts available at the test location, identify any anticipated repair procedures, etc. Having a plan in place prior to the initiation of flight test will allow the test team to anticipate and resolve potential issues without significant delays to the flight test program. Some recent envelope expansion projects have applied this lesson with improved results.

In general, there is a lack of awareness within the pilot community (both flight test and operational) regarding how airframe fatigue life is calculated. As a result, pilots tend to fly the aircraft without any regard to the potential impact on fatigue life. A notation in the flight manual stating that the airframe is not designed for repeated and prolonged exposure to the edges of the flight envelope may help to inform the pilot community of the potential fatigue issues without overly restricting the envelope.

SUPPLEMENTAL DISCUSSION: Examples of flight test fatigue failures due to flight test exposure beyond operational design usage.

Cracked Fuselage Panels on Legacy Tactical Fighter Aircraft:

During a test program on a legacy fighter platform, cracks were discovered on several fuselage panels following a flutter and flying qualities envelope expansion flight. The flight consisted of several test points at high Mach numbers and high dynamic pressures resulting in an accumulation of 13.6 minutes of effective time (at an equivalent dynamic pressure). In total, the aircraft had accumulated 75 minutes of effective time (at an equivalent dynamic pressure) in pursuit of envelope expansion testing. The cracks on the panels were attributed to sonic fatigue following extended exposure to supersonic conditions. Although these conditions were within the flight envelope, extended time at high Mach numbers and high dynamic pressure conditions were considered beyond the design envelope (sonic fatigue failure could be expected in as little as 9 minutes at high dynamic pressure conditions).

Pressure Relief Valve Failure on High Performance Fighter Aircraft:

During envelope expansion testing on a high performance fighter aircraft, there were multiple failures of a pressure relief valve (PRV) that was located in a weapon bay. The PRV failure was associated with a sharp rise in polyalphaolefin (PAO) reservoir pressure resulting in the draining of the PAO reservoir. All the failures occurred during testing at high Mach numbers and high dynamic pressures with the weapon bay doors open. In a comparison of the valves used in flight test with the valves used in qualification tests, the flight test valve displayed significantly more wear in $1/10^{th}$ of one lifetime (based on accrued flight hours) than the qualification test unit did at the completion of qualification testing. This discrepancy can be contributed to increased flight test exposure (versus design usage) and an inaccurate vibration environment used for the qualification tests.

Configurable Rail Launcher (CRL) Arm Cracks on High Performance Fighter Aircraft:

A post-flight inspection revealed a crack in the forward CRL arm following an envelope expansion test in the low supersonic region. The CRL is a trapeze launcher consisting of a forward arm, aft arm, hydraulic actuator, and rail launcher. The probable cause of the crack was determined to be sonic fatigue induced by the high acoustic environment in the weapon bay. Although there was a program desire to demonstrate the capability to launch missiles out to full envelope, the weapon bay design environment contained no design usage at the higher dynamic pressures allowed by the flight envelope. However in pursuit of clearing the full envelope, the amount of test time spent in the higher dynamic pressure environment (with the weapon bay doors open) leading up to this failure was ~51 minutes (~15.8 minutes on the same CRL arm where the failure occurred).

Trailing Edge Flaperon Failures on a Legacy High Performance Fighter Aircraft:

During loads flight testing, there were two separate failures of the trailing edge flaperon within approximately 18 months of testing. A subsequent investigation discovered that the flaperon design was based on the assumption of only two encounters of limit load conditions within the 8,000 hour airframe life. In contrast, during a single loads test mission, the flaperons were intentionally exposed to 100 percent Design Limit Load (DLL) conditions over 10 times in pursuit of data to support certification efforts. A comparison of flight test exposure versus operational usage revealed that 86 percent of the loads test points occurred in a region of the flight envelope where the operational fleet spends only 4 percent of its usage. Furthermore, 52 percent of the loads test points occurred in a region of the flight envelope where the operational fleet spends only 0.1 percent of its usage. Table 1 summarizes typical fleet service peacetime usage flight time.

Table 1 Peacetime Usage Flight Time

Flight Envelope as a Function of % DLL on the Flaperon	Hours / 8,000 Hour Life	% of Operational Flight Time Spent in Envelope
<60	7,680	96.0
<85	312	3.9
<100	8	0.1

Cracks in Primary Structure on a Legacy High Performance Aircraft:

Following extensive flutter and loads testing on the primary structures test aircraft, severe cracks were discovered in the aircraft backbone, the main spar of the wing, and other structural components of the wing and fuselage. As a result, the aircraft was grounded (later given a one-time flight authorization with a highly restricted flight envelope in order to return the aircraft to the contractor location).

CA11: GOVERNMENT ACCESS TO FUNDAMENTAL AIRCRAFT MODELS

LESSON LEARNED: When government flight testers are denied access to fundamental aircraft models for relevant technical disciplines, the ability to conduct safe, efficient, and effective flight testing is severely impacted. Given the contractual and legal implications, there may be no easy solution to this problem. The good news is that Program Offices are becoming more aware of this issue and have been more receptive to creative solutions than in the past. Methods are suggested that may help improve the overall situation for government flight testers while also protecting important intellectual property for the contractor.

PROBLEM: A very common theme has occurred on major EMD projects when the contractor stamps "Proprietary" on almost everything they produce or claims that every model is "intellectual property" and the government at times has been soft on challenging those assertions. Examples have included cases where the content had clearly been in the public domain, or when the content was clearly obtained from the government in the first place, or when the development of models had been paid for by taxpayer dollars. When the government does not establish clear contractual ground rules regarding access to fundamental aircraft models the results can have both near term and long term impacts. One of the main negative impacts of this trend has been that government flight testers were forced to accomplish their jobs while being relatively "blind" compared to the contractor.

DISCUSSION: For the purposes of this Lesson Learned, fundamental aircraft models are defined as those digital models that are designed to replicate the function and performance of the aircraft and its systems such as control laws, aerodynamics, engine, etc.

The heart of the historical problem is that the contractor is under no binding obligation to provide models if they are not specified as deliverables in contract documents. Even if that situation occurs on future projects, it does not mean that it is a hopeless situation, it just means that it could be far more difficult to obtain useable results. It has been common for the Program Office to define contractual deliverables that require the contractor to provide fundamental aircraft models to the government "at the end of EMD" or at "Functional Configuration Audit". Providing those models at such a late point in time may be sufficient to meet the long term government interests with respect to ongoing sustainment efforts, but does no good at all to help government flight testers participate in EMD flight test efforts. The wording of contract documents would be much more useful if flight testers are involved early enough to influence the statements for deliverables as recommended in the Appropriate Action section.

A common misconception is that the terms "Proprietary" or "intellectual property" mean that no one in the government can have access to those products. "Proprietary" or "intellectual property" does mean that the government needs to protect that information in an appropriate manner (i.e. restricting use to government agencies only). Some contractors have become very skilled at using the terms "Proprietary" or "intellectual property" as a shield just by implying that they "cannot provide access without the lawyers getting involved". The implied threat of legal entanglements is often sufficient to convince most technical or administrative people to cave-in and rescind requests for that product. This recurring situation could be avoided by clearly established ground rules in contractual documents regarding government purpose rights, or if already too late to influence the contract language… the ground rules and procedures could be in some other form that all participants agree to follow.

When past flight test projects have been clearly established as a genuine team of contractor and government flight testers, all stakeholders have had reasonable access to the fundamental aircraft models that provided a clear understanding of the system under test. This has often been due to the willingness of the contractor to participate on that kind of genuine team, not because of well-written contractual language. That clear understanding has enabled improved overall test planning, adequate preparations for each test mission, more confident interpretation of unexpected results during test execution, and enhanced post-test investigations. Unfortunately, most of the successful examples from the past have occurred as a result of the test team embarking on multi-year negotiations with the contractor and program office long after the contract was written. As a result of those lengthy negotiations, the test team was eventually able to convince the contractor and program office that sharing of fundamental aircraft models was in their best interest and should be interpreted as "in-scope". Ideally future successful examples will occur if conscious programmatic policies are imbedded in contractual language that give government testers access to aircraft models throughout the program.

When previous flight test projects were established in a way that prohibited a large portion of the test team from having access to those fundamental models, the flight test effort was not as efficient as it could have been, was more likely to miss characteristics that degraded operational capability, and has contributed to mishaps that were highly disruptive to the overall schedule. Sometimes the situation was purely the result of the type of acquisition strategy such as Total System Performance Responsibility (TSPR) or adaptation of a civilian aircraft to military uses. When a given contract explicitly identifies the contractor as having total control over the flight test effort, there may not be much that can be done to improve the situation for government flight test participants. In those cases the best hope for the test team is to convince the contractor and program office of the benefits obtained from model sharing.

When the government flight test organization is explicitly identified in the contract as the Lead Developmental Test Organization (LDTO), it can be a great help to legitimize the need for access to fundamental models, but may not help solve issues related to proprietary and/or intellectual property. Perhaps the best way to avoid these issues is to have the contract written to explicitly enable model sharing along with highlighting that government personnel are already under non-disclosure requirements.

APPROPRIATE ACTION: In an ideal world, flight testers should attempt to get the SOW or other contractual documents to explicitly require sharing of fundamental aircraft models with the government flight test team. The wording of contract documents would be much more useful if flight testers are involved early enough to influence the statements for deliverables to be something like, "Preliminary versions of the fundamental aircraft models will be delivered one year prior to first flight and will be updated regularly to accurately replicate the function and performance of the aircraft and its systems when testing shows discrepancies in the models and their related predictions. Final models will be provided at the end of EMD". Obviously that will only work if the government flight testers are involved early enough and the program office is supportive.

The biggest problems have occurred when the program office and/or contractor simply do not understand the need for government flight testers to have access to those models. Convincing them of the actual situation should not be relegated to negotiations by working level engineers, but needs to be aggressively supported at every level of test management. This has not been easy in the past but there are a number of ongoing higher level acquisition initiatives that test management could and should use as leverage to make progress in this area.

One initiative has been labeled "Digital Thread" (*The Air Force Digital Thread/Digital Twin - Life Cycle Integration and Use of Computational and Experimental Knowledge*, AIAA 2016-0897, reference 4), and has been championed by Dr. Edward Kraft, former senior Technical Advisor at AEDC. The overall goal of Digital Thread is to "reduce the length of the developmental test (DT) campaign". The scope of Digital Thread includes wind tunnel testing and Computational Fluid Dynamics and is therefore broader than the more limited applications needed by government flight testers. However, the availability of hi-fidelity aircraft models to enhance the ability of government personnel to support flight testing is a direct part of the Digital Thread concept.

Another high level acquisition system initiative has been labeled "Own the Technical Baseline" (*Owning the Technical Baseline for Acquisition Programs in the U.S. Air Force*, ISBN 978-0-309-37431-6, reference 5) and has been championed by Dr. William LaPlante, former Assistant Secretary of the Air Force for Acquisition. Owning the Technical Baseline means that the government program team, can make proper decisions to achieve successful acquisition outcomes. To accomplish that requires "deep understanding of system and sub-system designs and architectures" and includes availability of the models. The concepts associated with Owning the Technical Baseline tend to be from a program office perspective, but the flight test perspective can also be supported as a logical extension of those concepts. Program office access to the fundamental models is an important stepping stone to achieve flight tester access to those models.

"Digital Thread" and "Owning the Technical Baseline" both share the concept of a collaborative relationship between the government and the prime contractor. That collaborative relationship is essential regardless of the label that ends up being applied to the guiding concepts.

Another higher level process that has been influencing recent projects is the necessity for "vetting" of key models that will be used for requirement verification, capability verification, flight manual verification, etc. That vetting process has become more common in RFPs and may apply even if the models are considered proprietary. That vetting process may consist of Verification and Validation, Accreditation, or both. Portions of those processes may simply be a matter of providing adequate contractor documentation, but some portions may require more government oversight before the models are accepted for the intended purpose(s). Flight testers will need to understand how they fit within that higher level process for that particular project.

If flight testers are not involved early enough to influence the prime contractor's SOW or other contractual documents, it does not mean all hope of obtaining government access to models should be abandoned. It does mean that the process will be trickier and take longer. In that situation the first step is to obtain support from as many program office levels as practical. Obtaining support from the program office discipline level may sometimes be sufficient, but other situations may require additional support from program office management levels up to the Chief Engineer or higher.

The process of obtaining program office support will normally depend on rational and understandable explanations for WHY flight testers need access to those models (see Supplemental Discussion for examples). Examples of successful arguments that have been used in these situations were based on the fundamental premise that flight testers will always need that ability to accomplish their jobs safely, efficiently, and effectively. It may be possible to obtain limited access to the necessary models by stressing that without access to those models safety, efficiency, and effectiveness will all be degraded. This process will also require compromises on the part of government flight testers given that the contractor is being asked to provide something not explicitly required by contractual language (see Supplemental Discussion for examples of compromises that have been used successfully in the past).

Although it is highly logical to obtain program office support prior to setting out on a path to obtain contractor support, that may not always be practical depending on the individuals in various positions at the program office. Some program office individuals may have a lot of authority but not have directly relevant experience to help them understand the need. Others may have preconceived notions from prior projects that dominate decision making. In those situations a reasonably successful solution can still be achieved by working directly with contractor counterparts. Great care must be taken to avoid an implication of giving contractual direction. The contractor must understand it is in their own best interests to share the models. Like the program office, the contractor will need to understand WHY government flight testers need the models (refer to Supplemental Discussion for examples). It will also be very important to understand the contractor point of view and their concerns regarding use and protection of the models (refer once again to Supplemental Discussion for examples of compromises). Although some progress can be made by obtaining support from the contractor test organization, eventually the design team will also need to be convinced, since they tend to be the people who create, manage, and update the models.

The contractor could be provided a statement either in the contract or some other formal program documentation that explains that all government personnel working the program are bound by non-disclosure requirements. Many contractor personnel at the technical level may not be familiar with that constraint, which leads to their reluctance to share. Part of that statement can include the fact that only the program office can release products outside project channels (a restriction that is already quite common). In addition, some contractors have had no problem providing content or models to the program office, but were very uneasy about providing those same products to ANY other government organization. In those situations, some projects have successfully maneuvered through Proprietary or intellectual property issues by developing an agreement where the contractor will provide models to the program office, which then has the understanding and authority to decide if any other government agencies need access to those models. There have also been cases where the contractor was convinced their intellectual property rights were recognized and would be protected so they shared the info.

Another concept that has worked well on some past projects is to form a Modeling and Simulation Working Group or a Model Update Team. The best results have been achieved when those groups or teams operated under a formal Charter or Memorandum of Agreement (MOA). The Charter/MOA legitimized the activities of those groups and provided a solid foundation for ongoing collaboration, even if it was too late to have a direct impact on contractual language.

In general, it will be in the best interest of future flight testers to become more familiar with the concepts associated with “Data Rights”. This may not be a “topic of choice” for those flight testers (who would rather get more directly into the flight test business at hand), but may be necessary to enable future flight testers on that project to be in a much more solid position for accomplishing their jobs. Prior to contract award, those early testers may be able to influence the “Data Rights” contract language in very important ways (i.e. getting access to specific types of models, or getting that access earlier than originally envisioned by the program office). After contract award, the options will be much more limited and will require the cooperation of the contractor. In order to obtain that cooperation it may be necessary to coordinate additional complicated agreements given that the “Data Rights” aspects may not have been adequately captured in contractual documents.

Even in situations where the contractor has no philosophical objection to model sharing, they may have strong concerns about the extra workload required on their part to bring government engineers up to speed on how to properly use those models. Models for modern aircraft tend to be very complex. Improperly configuring the model, not knowing how to properly establish a trim condition, not understanding the true meaning of similar sounding terms, using obsolete versions, and many other aspects can all lead to wrong or misleading answers. Contractors are fully aware of how complex those models are and how easy it is to arrive at bogus conclusions. They will be understandably concerned about signing up for an open-ended task to train government engineers, especially when they aren’t being paid for that task.

The most obvious solution is to ensure that training government engineers on the use of models is specified in contractual language. That language will need to provide a clear indication that the task is not open-ended. One approach has been to specify that an initial cadre of government engineers would be trained by the contractor, but then that initial cadre would train subsequent engineers. It may also be important to specify that the initial cadre would consist of both program office and flight test engineers. Even if it is too late to influence the contract, this approach may still be workable if included in other agreements within discipline-oriented design or test working groups. Keeping the requested training from being open-ended may keep it below the contractor's threshold of concern and may be small enough to be interpreted as "in-scope", especially if supported by the program office.

There is another method that can be very helpful to ease contractor concerns about government engineers being incapable of successfully executing the models. That method has been to demonstrate extreme technical competence in various forums. This method can start in design reviews but has been demonstrated most effectively by participating in lab testing over a long time period (not just a "lookie-loo" session).

SUPPLEMENTAL DISCUSSION: List in the following are examples of WHY flight testers need access to fundamental aircraft models. These examples provide some specific ways that access to the models can improve overall flight test safety, efficiency, and effectiveness. In addition, flight tester access to the models will also improve their ability to accomplish the essential tasks for any flight test project of "Planning", "Executing", "Analyzing" and "Reporting".

Creation of a Genuine, Collaborative Flight Test Team:

This can be one of the most convincing and all-encompassing arguments. The most successful projects from the past have resulted when the government and contractor individuals on the test team were established as interchangeable for most of the routine flight test tasking. This does not mean that their roles will be identical since there will always be government tasking not accomplished by the contractor and vice versa. It does mean that many of the routine but essential tasks associated with flight testing can be shared based on individual priorities and availability on a given day. Routine tasks include examples such as: test plan and test card preparation, generation of pre-flight predictions, attending mission rehearsals in the simulator, attending pre-flight briefings, filling required roles in the mission control room, attending post-flight debriefings, and post-test data review.

Since that routine task cycle repeats many times during envelope expansion, the overall contractor workload can be greatly reduced when government testers can be part of the normal rotation for accomplishing those functions. Ideally that concept would be captured in contractual language describing the envisioned work-split. That will free up contractor personnel for other essential roles such as high level management briefings, preparing for delivery of new software and hardware capabilities, on-aircraft problem investigations, etc. A pattern established initially on many previous projects is that the onsite contractor personnel intend to be able to accomplish all tasks without government assistance. That level of effort by the contractor has sometimes been achievable early in the project. However, it has been very common that the contractor ability to accomplish all routine and essential tasks degrades rapidly as key personnel are called back to the home office and new personnel become hard to hire (see Lesson Learned "CA7: Plan for Shifting Work-Split within Test Team").

If the government members of the test team do not have access to the fundamental aircraft models, it may still be possible to achieve a genuine collaborative test team by splitting tasking as part of a compromise approach. However, it has been much more likely that the team will be dysfunctional with respect to important tasking. Some of the worst examples of dysfunctional test teams have occurred when neither the contractor nor government onsite test team members had access to the fundamental aircraft models. This has occurred when the onsite contractor test team members were viewed as merely "crank-turners" who were only responsible to accomplish test missions and ship out test data to the design organizations. The end result of that highly questionable philosophy was that no one in the mission control room was truly prepared with sufficient knowledge and understanding of the system to properly interpret normal versus unexpected test results (a VERY bad situation).

Some projects have had acceptable results when only the onsite contractor test team members have had access to the models for generating updated predictions or investigating unexpected results. In those situations the government test team members were highly constrained in their ability to be interchangeable partners in the flight test process. Those team members were unable to impart their unique expertise which would have enhanced test efficiency, provided more complete insights from the test results, and improved safety. The end result has been greatly increased workload for the local contractor engineers and a severe lack of real understanding for the government testers. Sometimes it has taken a mishap to highlight the undesirable consequences of this philosophy.

Generating Pre-Flight Predictions for Upcoming Test Missions:

One of the routine tasks for any envelope expansion test mission is the generation of pre-flight predictions that can be used in the mission control room to interpret test results as "expected" or "unexpected". This is an essential task that is often tied directly to termination or continuation criteria (i.e., terminate testing if a given value is exceeded, or continue testing if results are within predefined bounds). The initial plan proposed by a number of contractors has been for the design team in the home office (typically in a very different location than the test site) to generate those predictions for the entire test program and deliver that comprehensive package to the test team early in the project. This approach is often motivated by the very strong desire to maintain total control of the predicted data. That plan hardly ever survives intact for more than a few months given numerous factors that inevitably occur. Those factors include things like: aerodynamic models that need to be updated to reflect actual characteristics, revised software mechanizations to fix previous problems or implement new capability, ever-changing aircraft configurations, original predictions generated for a standard-day temperature profile that may not be achievable on a typically hot test day, etc.

Sometimes the contractor has re-run the entire predicted dataset multiple times in an attempt to keep up with all the changes. Eventually the design team has grown weary of that onerous, repetitive task and trained someone at the test site to assume that responsibility. When only a single person has been trusted with that role, that person inevitably becomes task saturated and becomes unavailable for other test support duties. It has been far better when this trend was recognized early on and multiple people at the test site were trained to properly run the simulations to obtain predictions. This approach has enabled the test team to generate those predictions on an "as needed" basis, usually within a few days or weeks of the test day when all of the variables have stabilized. An important additional benefit occurs when the local engineers generating the predictions understand the assumptions and constraints used, have an intimate understanding of the trends, and are better able to interpret unexpected results. The absolute best results have occurred when the engineers qualified to generate predictions have come from either the contractor or government members of the test team.

Better Decisions on Where to Focus Test Points:

A recurring problem has occurred on many test projects when the test point matrix was defined with too many points in the middle of the envelope and not enough near the edges or where the trends were more variable or uncertain (see Lesson Learned "FF6: Prudent Focus of Flight Test Points"). Access to the fundamental aircraft models can help everyone on the test team get a solid understanding of the expected trends. This will help define test points where they need to be instead of the common historical pattern of a somewhat-equally spaced set of points that cover the flight envelope. This includes accomplishing sensitivity studies on key parameters to guide decisions on where in the envelope it makes the most sense to locate test points. When the government flight testers have had access to the models they were in a much better position to contribute to a more efficient and effective test project as part of the team instead of acting as bystanders that tend to inflict unnecessary buildups as a safety precaution due to insufficient understanding.

Government Oversight of Entrance and Exit Criteria:

It is becoming more common for government flight testers to have a specific responsibility to independently assess readiness to enter a given test phase and to report on the key results from that test phase relative to predefined criteria. Those aspects may be incorrectly assessed if the test day conditions or procedures were not the same as the values intended by the criteria. This is especially true for assembling the components of mission performance that were obtained on different test days (e.g., takeoff, cruise, ingress, combat, egress, RTB, landing). Access to the aircraft models can allow the test results to be corrected to the intended conditions. Not surprisingly, some contractors may be reluctant to assist the government in accomplishing that oversight role. However, that role may be considered very important to the program office. It may be helpful to convince the contractor to share models for this purpose if they understand that the results would be obtained more quickly and accurately and therefore cause less unnecessary disruption.

Examples of compromises that may be needed to obtain reasonable access to aircraft models:

Executable Code Versus Source Code:

Government flight testers will always prefer having access to the source code for aircraft models because it will provide much greater insight into the design and enable flight test specific modifications locally that the contractor may not be paid to provide (i.e., corrections for test day conditions instead of standard day, etc.). However, it may not be practical to get the source code if contractual documents did not specify that as a deliverable. In that situation it may be best to work with program office and contractor counterparts to obtain access to models in an executable form, where the models can be run according to a predefined set of inputs and instructions, but with little or no insight into how the calculations are performed. Contractors tend to be more willing to share executable code because it provides better protection for aspects they consider to be intellectual property. Even when a given aspect of a model may not truly be intellectual property, the way that aspect is treated may boil down to what the program office is willing or not willing to support. In that case it would be much better to have access to executable code that can be used effectively versus having no access to aircraft models at all. Once the contractor sees that the government flight testers act as helpful and responsible partners using the executable form of models, they may be more willing to share source code even if not required to do so in the contract (given reasonable constraints as discussed below).

Written Agreements on How Models Will be Used and Protected:

One of the chief obstacles to contractor sharing of aircraft models (either executable or source code) has been their concern that those models will find their way into the hands of competitors. Another common contractor concern has been that the models will be used to "torpedo" their own efforts by taking issues directly up the chain to higher level program office management, but without adequate understanding of how the models need to be applied to get reliable results. A similar contractor concern is that disagreements on interpretation of results from models will lead to schedule delays whose main impact will be making it harder for the contractor to meet incentive requirements.

One method that has been used to address contractor concerns has been a written agreement that spells out the ground rules for how those models will be used. Agreements of this type do not need to be blown out of proportion into a major contract modification or a battle of lawyers. Typically, this should be a major function of the ITT. The agreement can also be crafted within the bounds of normal policies for a flight test working group or product team. The focus should be on how the team intends to work together to accomplish common goals, not on contract renegotiations.

The main contractor concern regarding models getting to their competitors can typically be handled by an appropriately worded distribution statement. Statements of that type probably already exist and may already be a part of the documentation for the models. All the written agreement needs to do is reinforce the intent of using that distribution statement. From the government perspective, that distribution statement should not be taken lightly. That statement should be taken just as seriously as any government distribution statement on reports or other products dealing with classification or constraints on release to foreign entities.

The contractor concern regarding taking perceived issues directly up to the external management chain can be handled by including explicit statements in the agreement to coordinate with them first and to provide an opportunity to come up with a consolidated team position on the issue before elevating outside of the test organization. The goal of these statements should be to place reasonable bounds on government conclusions obtained from use of models, but without applying excessive constraints. It is reasonable to give the contractor advance notice of a given issue derived from use of models, give them time to respond, and provide a method for coming up with a team plan before elevating that issue. In almost all cases there should be no need for the issue to be elevated at all. In a few cases, the contractor may be in a situation that prevents them from taking action on an issue derived from use of models even when that action seems warranted to the government members of the team. The agreement should acknowledge that the government members of the team may have an obligation to elevate a given issue once all the documented steps have been taken.

Some of the additional compromises listed below may also be appropriate for a written agreement. These are not intended as an all-encompassing list and each project will need to strive for an agreement that works best for them.

Don't Ask for Anything that Doesn't Already Exist:

One of the most effective ways to ensure that a given contractor becomes completely unwilling to share aircraft models, is to insist on the addition of a bunch of new features or capabilities that haven't already been included for their own purposes. It would be best for the government flight testers to adapt to whatever format, feature set, platform, etc. that the contractor already has in hand. That may not be considered optimal, but is the best way to get something useable in the first place. Once the models are being used effectively, and a satisfactory working relationship has been established, it may be easier to get a few extra features added if not too onerous for the contractor developers. However, in general it would probably be best to avoid requests like that. Another BIG advantage of only asking for products that the contractor will already be producing is that it will become much easier to get regular updates (as mentioned below).

Don't Ask for Every Engineering Version of the Models:

Contractors will typically produce many, many versions of the various aircraft models for every version that is representative of a flight test aircraft. Flight testers do not need all of the intermediate versions that the contractor uses internally for trade studies or ongoing development activities. Limiting the scope of model updates to be representative of the flight test aircraft is a good way to keep the contractor workload reasonable, and is much easier to justify to managers at the program office. The request will also be easier to justify if it only includes versions that will already be available within the normal contractor tasking (as mentioned above).

Don't Ask for Internal Analysis Tools that Have Been Developed At Contractor Expense:

One of the additional reasons contractors balk at model sharing, is that they often package those models internally with very complex analysis tools which are legitimately viewed as intellectual property. Those analysis tools have often been developed at contractor expense over many decades and on many different aircraft projects. Examples of those tools include: Linear and non-linear analysis of control systems, frequency domain analyses, parameter identification… and many more. The contractor provides those products as a package to their own engineers because they don't see the analysis task as being all that different than the need for the models. It's all part of the normal design tasking and sequencing. In general, government flight testers should already have access to their own similar analysis tools and should have no need for the contractor-developed tools.

TOPICS TO COORDINATE PRIOR TO FIRST FLIGHT

These lessons learned are intended to cover the time period in between contract award up to first flight. On some projects it may be appropriate to capture the top level aspects of these lessons in contractual documents but in general these topics are at a more detailed level.

The beginning of this period is a fairly discrete event and the specific wording of the contract will considerably limit the ability of flight testers to impact the process after that event. However, both the Program Office and the contractor will be motivated to be efficient, safe, and effective during envelope expansion so it may still be possible to influence the process if concepts are presented in an appropriate context. As any project gets closer to first flight it will become much harder to influence any process related to these topics. Therefore the “First Flight” lessons learned should be addressed soon after contract award and not delayed until almost at first flight.

Even after the contract has been signed, it can still be possible to negotiate useful agreements with the contractor, especially if support is first obtained from the program office. Sometimes these agreements can be formulated as simple Working Group “Ground Rules”, or a Working Group “Concept of Operations”, or something similar. Those agreements should be presented as “within scope” of the contract and merely as refinements. It will be important to avoid the perception of “requirements creep”.

FF1: CONFIDENCE TESTING PRIOR TO DELIVERY OF NEW CAPABILITIES

LESSON LEARNED: The concept of Confidence Testing has been used effectively on multiple flight test projects. Confidence Testing normally requires a team of flight testers to evaluate a new software/hardware combination at the contractor's most relevant simulation facility prior to delivery of that new capability to flight test. The main goal of Confidence Testing has been to expose the new capabilities to the realistic conditions and procedures that will soon be experienced during flight tests. An important bonus that also results from Confidence Testing has been that the flight test team has come away with an improved understanding of the new capabilities and much better preparation for the upcoming tests. Although the specific terminology and implementation has varied from project-to-project, the term "Confidence Testing" is used in this Lesson Learned because it provides a reasonably good description of the basic idea: increase confidence that the new software/hardware combination is actually ready for flight testing.

PROBLEM: On many flight test projects there has typically been an enormous amount of pressure to deliver new capabilities to flight testing "as soon as possible". While that is a worthy goal, it has often been achieved by delivering products that are not actually ready for flight testing. This has been a particular problem when award fee metrics have been tied to deliveries. The award fee language often only specifies the timing of the delivery, not that it must function as necessary to be effective during flight tests. A reasonable check and balance is needed to assess the actual readiness of the capability. Confidence Testing has served that purpose very effectively.

DISCUSSION: In practice, Confidence Testing has tended to focus on the software associated with new or updated OFPs more than hardware updates. However, Confidence Testing of some hardware updates has also been accomplished as long as the simulation implementation was affordable and practical. Typically these hardware-oriented tests were not accomplished on an iron bird (with full scale hydraulic systems and real flight control actuators) because iron bird simulations tend to be very expensive and difficult to operate. The most productive Confidence Tests have been accomplished on a piloted simulator with hi-fidelity models and with actual aircraft computer hardware and software in the loop (i.e., with software models of computer hardware components minimized). The aircraft computer hardware components do not necessarily need to be flight qualified (i.e., those particular units may not have passed environmental tests, electro-magnetic interference (EMI) tests, or other steps in the flight qualification process).

The end-to-end process for designing and verifying new products as implemented by most aircraft contractors has been incredibly methodical and rigorous, especially for safety-of-flight systems such as flight controls and propulsion. The overall success of the development effort depends on maintaining the discipline of the design process and the subsequent V&V process throughout the program. Nevertheless, in the early days of any development effort there are always growing pains as the entire team establishes a rhythm for producing new products. The first few software/hardware combinations are given a massive amount of attention in preparation for first flight and initial envelope expansion. However, at that early point in development the products are still very new and prone to unexpected results. Even as the system under test matures and the design/ V&V process becomes more stable, the pressures to "simplify" the end-to-end process tend to increase in order to meet ever more aggressive delivery schedules.

In some cases, new products have been delivered to the flight test community when those products were simply not ready by any interpretation. More typically, new products have been delivered to flight test with a number of unexpected issues, even though those issues may not have been considered "show-stoppers". A development effort is highly disrupted when very expensive flight test time is used to discover those unexpected issues when those same issues would have been easily observed during Confidence Testing. Unfortunately, Confidence Testing is a step that is frequently skipped in an effort to reduce short term monetary and schedule impacts without truly considering the broader implications. Confidence Testing provides the flight test team with an explicit opportunity to highlight the flight test impacts of various issues in situations where the Program Office is considering granting a waiver to delay fixes for those issues.

It is ironic that one of the strengths of the typical design/ V&V process is also one of the reasons that unexpected issues creep into flight test. The strength is that the design and V&V process is almost always requirements-based. The Systems Engineering concept of top-down requirements definition followed by bottom-up verification testing has been a very powerful process to ensure the end user gets a product that will do the intended job. However, many of those requirements are aimed at providing an Initial Operational Capability that is usually years away and not directly relevant to the immediate future of flight test. A fundamental question often overlooked in the rush to meet a delivery date is, "Can the new product execute as necessary over the next few months of flight testing?" It doesn't matter if the end-product will eventually be able to deliver cargo, fuel, or bombs as intended, if it can't properly accomplish the test points planned for the next dozen or so flight test missions. Confidence Testing fills the gap by exposing the new product to the essential test conditions and procedures planned for upcoming test missions and by providing an evaluation from the point of view of the same team that will be accomplishing those missions.

Typically, program office and design team leadership are not predisposed to support the concept of Confidence Testing and require a lot of convincing. Confidence Testing is initially perceived as just another delay in getting the new product to the test site. The best way to convince doubters is to show the benefits in action on their project, while also showing that perceived delays can be minimized. This may require "baby steps" in which the test team participates in the evolution of normal design-oriented simulator sessions. If that process is accomplished with a positive approach and aimed at genuine product improvement (i.e., not just nit-picking), most managers will see the value.

On some projects, the same managers that had vigorously opposed the concept became proponents. On those projects Confidence Testing (in some form) became both an accepted and expected part of the overall process. A successful Confidence Test came to be perceived as a "graduation exercise" to show internal and external overseers that all appropriate steps had been taken and passed. In the cases where "show-stopper" issues were uncovered during Confidence Testing, most management accepted that it was better to discover those issues prior to expensive flight testing. Even if none of the issues discovered were considered "show-stoppers" most management acknowledged the benefit of being able to start work on fixes sooner. It provided the test team with knowledge to determine the consequences of trying to test with these issues, assess the tradeoffs, and identify any practical workarounds prior to expensive flight testing.

The timing of Confidence Testing should be of paramount importance. The ideal timing is soon after normal V&V testing has been completed. On some projects the new product has already been certified for flight test. On other projects, Confidence Testing by the flight test team has been considered part of the final certification process. In typical development projects, Confidence Testing is often accomplished one or two weeks before the new product begins flight testing. Any sooner than that runs the risk that ongoing changes will be implemented after Confidence Testing and new issues might be missed. If Confidence Testing occurs within just a few days of starting flight tests on the new product, that doesn't leave much time to react to unexpected results.

<u>APPROPRIATE ACTION</u>: The flight test team (ideally including engineers, test pilots, and management from both government and contractor) should champion the concept of Confidence Testing long before flight testing begins. This can be treated as just a normal part of the flight test preparation process and possibly included in the CTF Charter. The program office and contractor design team leadership will need to understand the concept and provide an appropriate level of support. That support normally entails providing the flight test team with access to a piloted simulator with hi-fidelity and preferably with actual aircraft computer hardware and software in the loop and integrated in a way that reflects the test aircraft. The communication between hardware components should replicate the aircraft configuration as closely as practical (e.g., using the same network architecture). In order to function best for Confidence Testing, that simulator will also need to be configured with the appropriate capabilities to reflect off-standard-day flight test conditions, any required flight test aids, and a reasonable set of control room displays. Those capabilities can be relatively easy to incorporate if the coordination process begins early enough, but may be prohibitive if proposed just before testing begins. The flight test team will also need to negotiate the most timely point in the process to accomplish Confidence Testing.

SUPPLEMENTAL DISCUSSION: Confidence Testing Do's and Don'ts:

- DO NOT get hung up on the term "Confidence Testing." If the individuals on a given project prefer a different name… use it. It is far more important that this generic type of test occur similar to the concepts outlined in this Lesson Learned, whereas the specific name used for that testing is basically irrelevant. Some alternative terms that have been used include: pilot confidence testing, flight test rehearsals, test team Simulator Evaluations, etc. Each project will need to tailor the exact scope, coverage, and process.
- DO NOT allocate excessive time for Confidence Testing. A day or less should be sufficient for small capability increments. If the new product is a major upgrade with numerous capability improvements, adequate Confidence Testing can probably still be accomplished within two or three days.
- DO NOT repeat requirements-based testing from V&V. If the new product has been certified as having already passed the V&V process there is no reason to duplicate that effort. The focus should be on flight test specific testing that may not have been covered directly during V&V.
- DO NOT confuse Confidence Testing with normal test team mission rehearsals using a simulation facility located at the test site. Normal test team mission rehearsals are accomplished using a lower fidelity simulator that is primarily software based. Mission rehearsals are typically focused on an individual flight test mission (or small groups of related missions). By contrast, Confidence Testing tends to encompass a much larger "block" of flight test missions that can be accomplished after delivery of the new capability.
- DO accomplish Confidence Testing in the most representative ground-based simulator available. Typically this requires a hardware-in-the-loop simulator but does not require an iron bird. It is important that Confidence Testing be accomplished in a simulator that includes as many actual flight hardware and software computer assets as practical communicating over MUX bus architectures that represent the flight test vehicle. The specific timing of communications between each of those components can be one of the most important aspects to include in the simulator configuration. It is not necessary to use an iron bird because the expense and difficulty of maintaining and preparing all of those hydraulic and actuator components are not worth the extra fidelity. If a motion-based simulator cockpit can be connected to the hardware-in-loop simulator components, it might be worth considering as an option but should not be considered an essential requirement. In general, motion-based simulators add a limited amount of realism but are often not worth the additional complexity and expense. If the motion-based components break down during Confidence Testing, the tests should continue without motion since the most important goals can still be accomplished.
- DO require that Confidence Testing be accomplished by a small team from the flight test community (not just the designers). Flight testers will understand the upcoming test conditions and procedures and have a direct need to observe and understand any unexpected behaviors. For small

capability improvements the flight test team can consist of one or two test pilots and one or two flight test engineers. Ideally, a test conductor will also be present, especially if that same individual will also be conducting the first few test missions with the new capabilities. Major capability upgrades may require a larger team including three or four test pilots and a reasonable cadre of flight test engineers and test conductors given ongoing activities at the test site. It is NOT necessary for the test team to be only from government organizations. In fact, the best results are obtained when the Confidence Test team consists of both contractor and government flight testers. It is not reasonable to expect a single pilot/engineer team to accomplish simulator testing for 8 to 12 hours straight. The quality of their evaluation will degrade rapidly after 3 to 4 hours. For example, it has been effective to use one pilot/test conductor/engineer team for morning sessions and a different team for afternoon or evening sessions. This keeps each team operating at peak efficiency, improves scheduling options for personnel, and also provides for a broader spectrum of viewpoints (e.g., different experience levels on the project, OT and DT personnel, contractor and government, etc.).

- DO accomplish test points that reflect the upcoming test missions. This does not mean that every test point for the next 30 flights should be included. The overall scope should include a selection of the most important or challenging test points that will be flight tested with that software/hardware combination, but need not include every buildup point. If the upcoming flight tests will be accomplished during the summer at a desert location, the simulator should be configured at hot day test conditions, not standard day (and possibly also include significant turbulence and crosswind levels). If the upcoming flight tests will be accomplished primarily at heavy weight, Confidence Testing should focus on that weight. If the upcoming flight tests will focus on a particular discipline or portion of the flight envelope, Confidence Testing should focus on those aspects. The appropriate flight control modes and aircraft configurations for upcoming tests should also be established and evaluated in the simulator.
- DO attempt to provide the simulator control room team with displays that are representative of what will be available at the test site. The layout of the simulated control room does not need to be identical to the test site but the basic information presented needs to be as similar as practical. This typically requires significant advance coordination between the test team and the people operating the simulator. The best results are obtained when the simulator control room area is designed from the earliest days to operate using the same display tools as at the flight test site (Interactive Analysis and Display System [IADS] has become a de-facto standard).

Examples of issues that have been spotted during Confidence Testing: (this is not intended as a comprehensive list but just to show the types of issues that tend to escape the typical design process)

- Some software functions were connected to the wrong version of altitude (i.e., an above ground level [AGL] altitude instead of a pressure altitude or system-based inertial altitude). This was uncovered by running the Confidence Test at the test site location instead of the simulation's default sea-level location.
- The flight control software was not properly using outside air temperature. This was uncovered by running the hardware in the loop simulation using an off-standard day temperature profile as needed for upcoming flight tests.
- Some data being passed over the MUX Bus used the wrong word locations due to changes in the MUX Interface Control Document (ICD) for flight test that weren't known to the design team. Uncovered using a hardware-in-the-loop simulator with all systems configured as planned for flight test.

FF2: FLIGHT TEST FLEXIBILITY

LESSON LEARNED: On large scale EMD envelope expansion flight test projects one of the most important concepts that must be built into early flight test planning is the flexibility to expand the envelope along more than just a single path. Unfortunately this tends to be the exact opposite of the typical single-thread integrated test plans that have been developed, particularly when multiple stake-holders need to approve that consolidated plan. This Lesson Learned complements "FF4: Multi-Discipline Envelope Expansion."

PROBLEM: The basic problem has been the extreme difficulty experienced when trying to get stake-holders from multiple organizations and technical disciplines to agree on a general set of envelope expansion ground rules that maximize flight test flexibility. It has been far more common for the various organizations/disciplines to insist on a single, highly intricate integrated test plan that must follow a precise sequence of test points and any deviations must be re-coordinated at very high levels. Each organization or discipline has had valid technical and safety concerns that need to be addressed, but the extra effort required to define flexible envelope expansion ground rules has been considered excessive during early test planning when those ground rules need to be established. The end-result has been weeks or months of delay to the flight test process when the single-thread plan becomes derailed by the inevitable unexpected problems and a new single thread plan has to be coordinated with all stake-holders.

DISCUSSION: It has been a common program requirement that all fundamental envelope expansion flight test plans (Flutter, Flying Qualities, Loads, Propulsion, Vibro-acoustics, etc.) must be approved by the final authority weeks or months prior to first flight. That requirement was appropriate to ensure envelope expansion could move out aggressively shortly after first flight. The test plans for each technical discipline have typically been developed using concepts that make sense for each individual discipline, but were established independent of the other disciplines. When it comes time to consolidate all of those individual discipline requirements into an integrated flight test plan, it has often been too late to establish flexible ground rules. Many of those disciplines have already put so much effort into developing their independent test plans that no one wants to compromise. Instead, a complex and convoluted sequence of very specific test points is defined that represent the inflexible requirements established by the combination of all disciplines. The resulting sequence can appear to be reasonably logical and well thought out, but the flight test organization becomes trapped into that single path.

Once serious envelope expansion flight testing begins, unexpected problems are encountered relevant to one or more disciplines. It is common for those problems to halt envelope expansion along the pre-approved path until a fix is provided, usually in an unplanned OFP update. It can take weeks or months for that unplanned OFP update to be delivered to the flight test organization. In the meantime, envelope expansion needs to continue since so many other test requirements are dependent on having a cleared envelope. When this occurs, it is usually fairly obvious what needs to happen to allow envelope expansion to continue on a different path such as: go to a different airspeed/altitude, test in a different configuration, backup in terms of dynamic pressure but start testing more aggressive maneuvers sooner than planned, etc. However, that obvious alternate path often takes many weeks to be re-coordinated and approved amongst multiple organizations/disciplines that all have competing priorities at that point in time.

The overall flight test envelope expansion process would be far more efficient if all stake-holders agreed up-front to a few fundamental ground rules that would allow for testing along alternate envelope expansion paths when the primary path becomes derailed. This would enable a much more agile test organization that can continue making worthwhile progress since the ground rules have been pre-approved by all stake-holders.

<u>APPROPRIATE ACTION</u>: It is unlikely that a flexible set of alternate envelope expansion test paths will be available unless the requirement to define reasonable ground rules is established very early during the overall test planning process. Flight test leadership needs to legislate that all stake-holders establish fundamental envelope expansion ground rules, NOT just a single, complex path through multi-discipline test matrices. The requirement for establishing these fundamental ground rules needs to be established by an advanced planning team long before the individual disciplines have already established their own detailed test concepts. Once representatives from each organization/discipline have agreed to the fundamental envelope expansion ground rules, a side benefit would also occur because it would make the process of developing the initial integrated test plans much easier.

Once the fundamental envelope expansion ground rules have been established they will need to be captured in relevant test and safety plans. Those ground rules will need to be understandable enough so that reviewers don't feel like they are signing a blank check. The wording will also need to be clear enough so that all concerned understand when a given alternative path is a normal part of flight test flexibility and when a change triggers the need for additional review or clearance.

<u>SUPPLEMENTAL DISCUSSION</u>: To be successful, flexible envelope expansion ground rules do not need to provide for every possible contingency. It would be sufficient if those ground rules capture a few logical decision paths that would probably be accomplished anyhow when the primary envelope expansion path becomes derailed for whatever reason. Examples of useful envelope expansion ground rules:

- When the primary Flutter envelope expansion path is halted, testing can continue on other paths as long as the dynamic pressure is at least XXX psf lower than previously successful Flutter flight conditions.
- When the primary Flying Qualities envelope expansion test path is halted, testing can continue on other paths as long as the Mach number is at least 0.X slower and the airspeed is at least XX knots slower than previously tested flight conditions.
- When the primary Loads envelope expansion test path is halted, testing can continue on other paths as long as the predicted loads for the intended maneuver are less than XX percent of design limit load.
- Testing in the Clean configuration does not have to be 100 percent complete to begin testing in other configurations. For example: testing with weapon bays open can commence as long as the flight conditions are less than previously cleared for the Clean configuration by XXX knots, 0.X Mach number, etc. A similar ground rule can be established for alternate flight control modes, control surface configurations, external store loadings, etc.

Obviously the actual envelope expansion ground rules for a given project will probably be somewhat more complicated than the examples listed above. Each discipline will insist on tight constraints in order to maintain a safe envelope expansion process. The essential point of this entire discussion is that the ground rules necessary to enable a few alternate flight test paths need to be predefined and acceptable to all participants. This will enable the test team to proceed smoothly without a lengthy re-coordination process. Since the test team usually includes representatives from all disciplines, they will still have an opportunity to object if the alternate test path poses some unforeseen difficulty.

Another aspect of providing alternate flight test paths includes the fact that the test aircraft must be capable of testing along those alternate paths when necessary. For example, testing with the weapon bays open would not make a viable alternate test path if the vibro-acoustics instrumentation was not ready when needed. This goes against the common principle of "just-in-time" functionality that seems to have driven many previous flight test programs. Instead, it would be better for the flight test program if those functionalities were delivered based on timing that reflects "as early as might be needed" versus the more typical "not functional until the last possible need date."

FF3: IN-FLIGHT SIMULATION PRIOR TO FIRST FLIGHT

LESSON LEARNED: In-flight simulation has been used effectively for risk reduction prior to first flight on new airframes. Examples have included the X-15, YF-16, AFTI F-16, B-2, F-22, and X-35. Significant control law changes have been implemented as a result of in-flight simulation and the results have helped improve overall confidence during the First Flight Readiness Review process.

PROBLEM: This Lesson Learned is aimed more at communicating a best practice than it is capturing a problem/solution. In-flight simulation may not be necessary on every new airframe, but past results have uncovered the need for changes that were implemented prior to first flight and may have helped avoid surprises during early flight test. Indicators that in-flight simulation may be warranted can include unusual design characteristics such as a unique control inceptor, an atypical design with the center of rotation forward of the pilot, or control law architectures that have not been attempted on past projects.

DISCUSSION: The intent of this Lesson Learned is to capture some of the tradeoffs that have been experienced during in-flight simulation on past projects so that future projects may be able to make more informed decisions and be better prepared if in-flight simulation is pursued.

In the context of this handbook, the main purpose for in-flight simulation has been as a confidence builder prior to first flight. Therefore the primary focus has tended to be on handling qualities during basic power approach landing tasks. However, a few other tasks with particular interest to specific projects have also been evaluated such as simulated aerial refueling and crosswind landings. In-flight simulation may be particularly important to reduce development risk for cutting edge designs such as when a legacy fighter was first to employ a force sensing side stick controller and fly by wire flight controls. There is no need to use in-flight simulation in an attempt to evaluate every operational task for a given platform. Even though in-flight simulation may be capable of providing reasonable matches across a broad spectrum of tasks and flight envelope, the focus should be on the main uncertainties associated with first flight and initial envelope expansion.

The real secret to in-flight simulation is preparation. If the designers and test team on a given project commit adequate data, engineering time, and management coordination then the results should be worthwhile. If in-flight simulation is treated as an after-thought with little preparation/coordination prior to the testing then the results will probably be much more problematic and controversial and therefore not worth the investment. Past projects have exhibited both ends of that spectrum and a few in-between.

Although in-flight simulation has improved over the years it remains somewhat of an art form. It can be quite tricky to make one aircraft fly like another. For example, it may be possible to produce an excellent match in pitch rate response while sacrificing the match for load factor response. That difference can be more significant in some piloting tasks than others. Similar issues can exist in the roll axis matches. Greater success has been achieved when the designers and in-flight simulation team agree ahead of time on how good those matches need to be in each axis.

Some in-flight simulator aircraft have special control surfaces (or special control laws using existing surfaces) to help improve matches with predicted behavior in terms of both rate and load factor. Typically, those additional capabilities have limited authority and may not be able to exactly match the intended response for all characteristics. It is important that project designers and test team understand those constraints and work closely with in-flight simulation personnel to minimize the impacts. At some point, the differences may be enough to negate any usefulness as a design tool or predictor of actual aircraft response. Making that decision would require a lot of good, solid engineering judgment from the collaboration of the test team and in-flight simulation team.

If the intended design used a unique control inceptor of some kind, it was important for the in-flight simulation aircraft to be modified to accept a representative version of that inceptor. The unique control inceptor was only required for the evaluation pilot position, not for the safety pilot.

APPROPRIATE ACTION: Flight testers on future projects for new airframes should ensure that a serious discussion of in-flight simulation takes place with the project office and contractor well ahead of first flight. The result should be a genuine evaluation of the pros and cons of in-flight simulation for that platform. In-flight simulation should not be dismissed out of hand because of the cost or other perceived issues. Any pre-conceived issues need to be properly weighed against the broader programmatic risks associated with not doing in-flight simulation. Even though in-flight simulation will undoubtedly be viewed as expensive, those costs can be far less than the programmatic risk of undiscovered issues getting into flight test on the much more expensive test aircraft, along with the related costs due to schedule delays to implement fixes that could have been implemented prior to first flight.

One of the biggest constraints for future projects may be the retirement of many old aircraft from the in-flight simulation fleet. The specific capabilities/authorities/features of the remaining in-flight simulation aircraft need to be understood and considered before making any decisions.

SUPPLEMENTAL DISCUSSION: One of the most significant side-benefits of in-flight simulation has been observed during the First Flight Readiness Review (FFRR) process. The “gray hairs” who are typically invited to participate in FFRRs are usually independent of the project and are well aware that optimistic predictions prior to first flight have proven to be well off-target on previous projects. In-flight simulation has helped convince those skeptics that all reasonable efforts have been taken to prepare for first flight.

There will probably always be proponents of ground-based motion simulators as an alternative to in-flight simulation. Motion simulators of various types have been used on many projects, but tend to have similar issues as in-flight simulation in terms of simultaneously matching rate and load factor responses. Motion simulators also have constraints in terms of authority and the maneuver duration in which the target characteristics can be matched reasonably well. For example, turbulence response has been replicated in motion simulators, but few pilots have thought that the result was a useful representation of actual atmospheric turbulence. However, those same pilots have pointed out that a given task was made more of an interesting challenge when simulated turbulence was introduced into the motion simulator, as opposed to the “glass-smooth ride” that was experienced without simulated turbulence. On some projects that had regular access to a hi-fidelity motion simulator, the decision was still made to pursue in-flight simulation in order to put the aircrew into a real flight environment with real boundary conditions for landing tasks (i.e., the ground) as opposed to projections on video screens.

In-flight simulator aircraft always have a safety pilot with the ability to take control using the baseline flight control system if unexpected handling qualities are encountered with the dynamics of the platform being simulated. In addition, in-flight simulator aircraft typically have automated “trip-off” criteria that will revert to the baseline aircraft if pre-defined limits are exceeded. Understanding of these implementations is critical to use of the system.

In addition to influencing some aspects of the basic control law design prior to first flight, another side benefit of in-flight simulation has been to highlight areas of uncertainty in the design. It is not unusual for designers to feel they have a rock-solid implementation going into the in-flight simulation process (usually based on a combination of the various numerical handling qualities criteria and fixed-base piloted simulators). After in-flight simulation, that confidence may not be quite so high and designers have been more willing to accept the possibility of wider uncertainties in terms of predicted aerodynamics along with the potential implications for closed loop handling qualities issues.

A related positive outcome has been when those uncertainties have led to genuine consideration applied to the nature of flight test aids (such as off-nominal gain variations). Having appropriately considered flight test aids designed into the first test article has enabled test teams to more quickly respond to unexpected results and continue envelope expansion while using a pre-designed flight test aid feature. Using flight test aids during flight test (as needed when unexpected results are encountered but a more formal update is months away) has also given designers more confidence that the gain changes or other feature changes that will be provided in more formal control law updates are appropriate. See Lesson Learned "CA5: Incorporate Flight Test Aids on Test Aircraft."

FF4: MULTI-DISCIPLINE ENVELOPE EXPANSION

LESSON LEARNED: On large test programs involving extensive envelope expansion testing (flying qualities, loads, flutter, vibroacoustics, weapon separation, etc.), the specific test points are typically planned within each individual discipline stovepipe. This isolated test planning has prohibited integration into a logical, cohesive envelope expansion plan. The end result was workable but highly inefficient, delaying achievement of program objectives. This Lesson Learned supplements "FF2: Flight Test Flexibility."

PROBLEM: Envelope expansion testing requires a build-up in flight conditions (airspeed, Mach number, dynamic pressure, altitude, etc.) and test maneuvers to safely test and expand the flight envelope. On most large EMD test programs the test planning for each discipline has been done within the discipline stovepipes resulting in test point flight conditions and maneuver types that do not correlate across disciplines. This stovepipe approach in the test planning process misses the opportunity for the disciplines to collaborate on test conditions, test maneuvers, and build-ups that would simultaneously satisfy each of their respective requirements and enable much more efficient flight testing.

DISCUSSION: The history of older test programs with extensive envelope expansion testing shows that the specific test points were typically planned and flown as dedicated single discipline test flights. This approach was an artifact of much simpler, more federated weapon systems where envelope expansion could be accomplished efficiently one discipline at a time. The complexity of modern highly integrated weapon systems prohibits using a single discipline approach. The more modern control law architectures are designed to provide better protection from exceeding structural limits. For example, control laws have features to limit roll acceleration in order to keep wing bending loads and torsional loads within acceptable boundaries. Another example occurs at the leading edge of envelope expansion (highest dynamic pressure tested to date) where instabilities can occur due to different mechanisms related to flutter margin, aeroservoelastic margin or rigid body dynamics. Therefore, each incremental step toward the envelope limit must involve a simultaneous assessment of the flying qualities, loads, flutter, etc., in order to progress to the following increment toward the limit.

Many of the disciplines involved in envelope expansion testing employ a build-up process that could enable collaboration on test conditions, test maneuvers and build-up approach that simultaneously meets their respective requirements. However, this has not occurred effectively because each discipline has established fundamental flight condition requirements independently. For example, the flying qualities discipline tends to base test conditions on wind tunnel testing. The loads discipline tends to base test conditions on the static predictive models whereas the flutter discipline tends to select test conditions based on dynamic predictive models. The primary impact on flight test is that each discipline selects test altitudes that do not correlate, forcing significant amounts of altitude changes.

Some of the test disciplines also have very unique test maneuvers/techniques that are necessary to meet their requirements. However, these maneuvers/techniques are typically conducted in a similar build-up approach to the other disciplines. These maneuvers/techniques can be integrated into a block of testing combined with other disciplines as long as test conditions correlate. The most significant aspect that has been missed on previous projects has been when test altitudes do not correlate across disciplines. Consequently the test team was encumbered with an intensely difficult task to define a test path that would be satisfactory for all design disciplines. None of the design disciplines were willing to compromise because they would miss getting data that matched their prediction conditions. The resulting flight test path inflicted very time consuming and fuel consuming changes in altitude from test point to test point.

APPROPRIATE ACTION: In the early planning stages of envelope expansion test programs, the flight test organization needs to actively advocate an approach to test planning that will lead to an efficient and effective Envelope Expansion Plan that integrates all relevant flight testing. This type of integrated plan is best initiated near the beginning of a development effort in order to establish basic concepts and principles for all technical disciplines. The Envelope Expansion Plan should be developed in parallel with the individual discipline-oriented test plans, not just assembled in an ad hoc fashion after the lower level plans have already been coordinated and approved at high levels. The people assigned to develop the Envelope Expansion Plan must have extensive cross-discipline visibility and understanding. Those people will also need a lot of support from senior flight test management because it will probably be necessary for individual technical disciplines to be directed to accept unattractive compromises. Once the Envelope Expansion Plan has been matured, it will undoubtedly need to go through a technical and safety review process (in addition to similar processes for each discipline-oriented test plan) and obtain management approval consistent with the assigned risk level for the combined testing.

The key developers of the Envelope Expansion plan will need to coordinate with the relevant technical disciplines and convince them to collaborate on test conditions, test maneuvers, and build-ups that would simultaneously satisfy each of their respective requirements. This would make the overall envelope expansion testing significantly more efficient and enable a much faster clearance of the flight envelope to allow other testing to commence. In particular, test altitudes need to correlate across test disciplines. Ideally Mach numbers and airspeeds would also correlate, but this is not as significant as test altitude.

There is another aspect of Multi-Discipline Envelope Expansion that should be given strong consideration long before flight testing, but is often given no advance consideration at all. That aspect is incremental envelope clearance. It should be self-evident that flight envelopes are not cleared all at once, but must be cleared in increments that represent the consolidated efforts of multiple technical disciplines.

Those incremental clearances are typically represented in a consolidated "Operating Limit" of some kind. Since these incremental envelopes tend to be well short of the intended full envelope, the Flight Manual is not the best place to capture the partial envelope definition. Each project/contractor uses a different mechanism for that purpose but the intent is usually very similar. Examples include: Aircraft Engine Operating Limits (AEOL), Flight Operations Limitations Document (FOLD), Aircraft Operating Limits (AOL), or simply Operating Limits (OL). Regardless of the name, those operating limits are usually initiated by each individual technical discipline and hopefully represent the least restrictive set of limits that can be provided by that discipline at that point in time. There has been nothing wrong with that part of the process… when each discipline applies an appropriate amount of energy and rigor to expanding those discipline-specific limits.

However, it has been very common for consolidated multi-discipline envelope clearances to be obtained somewhat randomly based on whatever collection of successful test points have been accomplished to date (combined with "carve-outs" for portions of the envelope where problems have been encountered). When each discipline has defined test points at inconsistent flight conditions, the process for defining a consolidated envelope clearance has become even more problematic and contributes to the resulting envelope being less useable than it could be. It would be far better if those clearances were goal-oriented with each discipline working together towards a common consolidated envelope that provides a meaningful payoff to enable other types of important testing (Mission Systems, Weapon Separations, Radar Cross-Section Testing, etc.). The more typical approach has proceeded along paths defined by the competing interests of each discipline. The result has often been a collection of very odd operating limits that don't support other test types very well at all.

Not much can be done about the need for envelope "carve-outs" when problems are discovered during flight test (other than reducing the length of time those constraints are in place as discussed in "FF5: Providing Focus on Fixing System Deficiencies Impacting Flight Test."). However, the same techniques described earlier in this Lesson Learned can have the side-benefit of improving the definition and usability of incremental envelope clearances.

It is also highly recommended that each project take a proactive approach to pre-define meaningful goals for incremental envelope clearances at important points in time. Examples include: the minimum envelope needed to begin Mission Systems flight testing, the partial envelope needed to accomplish the next sequence of Weapon Separation testing, the comprehensive envelope needed to accomplish the next series of Integrated Systems Evaluations). Each project will need to determine which ensuing test types most urgently need a cleared envelope, and define the objectives for each incremental envelope accordingly. There is no intent for those incremental envelope clearances to be targeted weekly or monthly. Something like 2-3 incremental envelope clearances per year seems like a reasonable rate, but the timing should be driven by real needs, not just to provide periodic updates.

Once a meaningful set of incremental envelope clearance objectives have been established, reporting on the accomplishment of those objectives should be built-in to the advertised plan for providing progress reports outside the project. Reporting that a pre-defined envelope has been cleared on schedule to open up another important type of testing will have MUCH more meaning than just reporting on how many test points have been accomplished and how many flight hours have been flown (see Lesson Learned "CA8: Tracking Flight Test Progress").

SUPPLEMENTAL DISCUSSION: The discipline-oriented stovepipe approach is typically the result of a lack of systems engineering type thinking relative to the test and evaluation, and/or driven by accounting processes to keep a clear accounting of expenditures and progress in accordance with a stove piped accounting structure. Conducting envelope expansion testing in flights dedicated to a single discipline requires each discipline to go back to essentially the same flight conditions and repeat many similar, if not the same, test maneuvers.

Integrated Test Blocks:

The Integrated Test Block (ITB) concept has worked very well as the key element of Envelope Expansion Plans on several big envelope expansion projects. The intent of ITBs is to group test points from individual technical disciplines at a single test condition in order to achieve a common goal. Each project will need to determine how to best define those goals and how to best group those test points.

One type of ITB that has become almost ubiquitous is the LEITB (See the Discussion paragraph under "CA10: Flight Test Fatigue Failures Due to Flight Test Exposure Beyond Operational Design Usage" section). The intent of an LEITB is to group a relatively small number of multi-discipline test points that all need to be accomplished the very first time a new flight condition is attempted. Typically this applies to first time testing at a higher Mach number or dynamic pressure. Although LEITBs are often associated with Flutter testing, that does not mean the very first test points at a new flight condition need to be exclusively Flutter test points. It has become common for the first test points to be focused on basic stability checks using "stick raps", "pulses", or "singlets" (minor variations with the same basic intent). Regardless of the name used, the goal is the same… to confirm adequate damping from relatively short, small amplitude excitations before proceeding to larger or more prolonged excitations or maneuvers. That type of small amplitude excitation has been very useful to simultaneously verify adequate damping for several disciplines (Flutter, Stability and Control, ASE, etc.). Once basic stability has been confirmed the test team can proceed with test points for the next most critical discipline, typically Flutter or Stability and Control, but sometimes Structural Loads, Propulsion, or some other discipline. Each project needs to figure out "Which discipline goes next?" and then "Which discipline after that?" until a given LEITB is complete.

An LEITB is generally accomplished at or near 1g. Gentle maneuvering is often accomplished in each axis to include small amplitude doublets, frequency sweeps, and mild bank-to-bank rolls. The last maneuver in an LEITB is often a gentle pitch axis maneuver to a benign g-level for that aircraft.

Once an LEITB at a new test condition is complete, the Envelope Expansion Plan should outline the next step. One obvious option would be to step to the next LEITB in a sequence. However, a long sequence of LEITBs near 1g should not be accomplished too far ahead of an Integrated Test Block that expands into more significant maneuvering such as pitch axis maneuvers up to moderate g for that aircraft, larger amplitude rolling maneuvers, etc. If a long sequence of LEITBs is accomplished without a reasonable amount of maneuvering expansion not too far behind, the test team could paint itself into a corner if a real instability is encountered, the Test Conductor calls "Knock-It-Off", and the pilot needs to quickly return to a lower airspeed. In that situation it may not be sufficient to simply retard the throttle to IDLE power since the aircraft may not decelerate quickly enough. It may be unwise for the pilot to deploy speed brakes unless that option has been cleared along with the previous LEITBs. Deploying speed brakes at much higher airspeeds than has been done previously would amount to unplanned envelope expansion and could just make a bad situation even worse. Besides going to IDLE power after a "Knock-It-Off" call, pilots are normally given the option to add load factor up to a benign g-level to increase drag and more quickly get to a lower airspeed. The g-level attained in previous LEITBs may or may not be sufficient for that emergency deceleration situation.

After the LEITBs, various projects have used different names for the next type of ITB. Examples include Maneuvering ITB, Follow-On ITB, Trailing ITB, and others. The specific name is not important as long as it is meaningful and useful for that project. Regardless of the name, the goal of those ITBs is to follow a logical path that enables safe and efficient envelope expansion for multiple disciplines to eventually clear the entire maneuvering envelope. Those ITBs are used to gather data to clear other dimensions such as load factor, rolling maneuvers, angle of attack, angle of sideslip, combined roll-pull and/or pull-roll maneuvers, etc. The testing along each of those dimensions must be planned with answers to the same questions mentioned earlier: "Which discipline goes next?" and then "Which discipline after that?"

Envelope Expansion Ground Rules:

As discussed in Lesson Learned "FF2: Flight Test Flexibility", it will be very important for the Envelope Expansion Plan to define more than just a single path through envelope expansion. Numerous types of problems can and will be encountered during envelope expansion that will quickly derail the most well thought out single path. Many of those problems will prohibit testing along a single path for many months until a root cause can be determined, and a fix designed and delivered. This does not mean that the Envelope Expansion Plan should include a dozen different paths. It would be OK for an Envelope Expansion Plan to define a "preferred path" since that might be the easiest way to get all stakeholders to agree on something definitive, but that should not be the ONLY path. The best approach would be for the Envelope Expansion Plan to first define the overall methodology, and then define the fundamental rule sets that can be used at any point along various paths. Those fundamental ground rules need to be logical and understandable to each discipline and to the entire chain of signature authorities all the way up to the highest approval level. The ground rules cannot be so vague that they appear to be a "blank check" to take ANY alternate path. Examples of fundamental envelope expansion ground rules (these must be highly customized for each project):

- The step size between LEITBs cannot be larger than XX knots, 0.X Mach number, or XXX psf dynamic pressure

- If the preferred path through an LEITB Mach number sequence at a given altitude cannot be continued, an alternate LEITB can be executed at a lower (or higher) altitude as long as Mach number is at least 0.X lower, airspeed is at least XX knots lower, and dynamic pressure is at least XXX psf lower.

- If a Maneuvering ITB is limited by a load factor restriction, the test points in that ITB can still be accomplished up to the restricted load factor value in order to obtain an incremental envelope clearance. Once the restriction is lifted, portions of that Maneuvering ITB will need to be repeated for final envelope clearance.

It will not be easy to get all stakeholders to agree to the same set of fundamental ground rules. However, there could be a huge payoff when unexpected problems are encountered and the test team has the pre-approved option to shift onto an alternate path that is also constrained by the same existing ground rules. The only alternative to a well-defined set of envelope expansion ground rules is to wait until problems are encountered and then react with a new round of coordination (usually taking weeks or months) to define another unique path that will probably last only a short time until another problem is encountered.

Flight Test Continuation Criteria (FTCC):

In addition to the fundamental ground rules used to make decisions on what path to take through multi-discipline test points, each discipline will need to establish pre-defined continuation criteria (a few specific examples are discussed in "FF8: Residual Oscillations" and "FF9: Rate Limited Oscillations"). Those continuation criteria need to be used to guide the interpretation of the results from each flight test maneuver. If the results are within those criteria, the discipline test team normally decides to proceed to the next test point. If the results are not within those criteria, the discipline test team is normally obligated to stop testing along that path for the duration of that flight (testing along other paths may continue). Once the discipline team has had time to review data after the flight, and sometimes coordinate with the related design organizations, they may decide to resume testing on the same path on a following flight.

In modern times, most experienced flight testers recognize the value in pre-defined FTCCs. They have accepted the premise that it is far better to hash out all the tradeoffs as a cohesive team long before seeing unexpected results in the mission control room, and then making critical decisions as individuals subject to high stress situations during a challenging test mission. This broad acceptance of the concept for FTCCs has not always been the case. In past times, many discipline engineers (either government or contractor) were highly resistant to "having their hands tied" by predefined criteria. They wanted to make all judgements based upon their own personal experience and philosophy. The problem with this approach has become obvious when the control room team supporting each mission had a wide range of experience (often the case) and differing test philosophies. Flight test history has repeatedly shown that it is unwise to make those critical, envelope expansion-related decisions "in the heat of the moment". The overall results have been much more successful when the discipline test team coordinated FTCCs with each other (tricky but achievable), and with the related design organizations (much harder… but worth the effort).

Well defined FTCCs help guide point-to-point envelope clearance decisions such as within an LEITB, from one LEITB to another, or from an LEITB to a Maneuvering ITB. The FTCCs are not typically used for flight-to-flight clearance decisions since those are usually made with more time available. The exception has been when one envelope expansion flight is conducted in the morning by one set of discipline team members and an afternoon flight along the same path is conducted by a different set of individuals. In those situations, well defined FTCCs have helped provide for a clean "handoff" between the two teams.

This does not mean that FTCCs need to cover every technical aspect in excruciating detail and therefore completely remove any real time judgment from the discipline test team. The goal of establishing worthwhile FTCCs should be to capture the outer bounds for the fundamental characteristics that NEED to be evaluated before proceeding to the next test point in a sequence. Typical examples have included things like: peak sideslip compared to predictions, maximum amplitude for residual oscillations (an indication of basic instability), measured loads on an important strain gauge (such as at the wing root), airspeed error for the indicated value feeding the flight control system compared to a truth source (pacer, test nose boom, etc.), magnitude of overshoot beyond an angle of attack limiter, and many others. The good news is that any new project should be able to review the FTCCs actually used on recent projects for a similar aircraft type and use those to determine what might work best for the new project.

The concept of FTCCs has worked best when not overly formalized. For example, it has been advisable to avoid including specific values as part of safety package minimizing procedures. It has become common for the safety package General Minimizing Considerations to include the requirements for the discipline test team to establish a predefined set of FTCCs prior to each test mission, and require the entire test team to abide by those FTCCs, but the specific values can be left to the discipline test team. This approach is appropriate because the specific values selected often depend on where the test condition is located within the overall flight envelope and may also be influenced by other factors (such as old vs. new software implementations).

Compatible Instrumentation:

There is a significant amount of overlap in the instrumentation required by the various disciplines involved in envelope expansion, and there are also instrumentation parameters that are unique to certain disciplines (i.e., structural loads data). Stove piped planning by discipline and by flight may preclude ensuring the test aircraft has the necessary instrumentation to conduct integrated multi-discipline envelope expansion testing. This happens when the required instrumentation is distributed among different test aircraft. Early planning with focus and intent on integrated multi-discipline envelope expansion testing would ensure the test aircraft were instrumented appropriately for this approach. A necessary part of the engineering discipline collaboration would be related to the instrumentation parameters required, along with the necessary conditioning, resolution and accuracy to meet the respective engineering requirements.

FF5: PROVIDING FOCUS ON FIXING SYSTEM DEFICIENCIES IMPACTING FLIGHT TEST

LESSON LEARNED: When flight testing discovers dozens or hundreds of problems that need to be addressed across multiple systems and subsystems, the process for prioritizing and fixing those problems has lacked focus on the consequences to flight test and degraded to an oversimplified tracking system. There has been a lack of balance between addressing important problems that need to be fixed for the production configuration while also providing timely fixes to problems with near-term flight test impacts. The unintended consequence has been the delay of resolution of problems which created inefficient progress through flight test, prolonged flight test schedules, and additional costs.

PROBLEM: On many large EMD-type development projects, decisions to fix (or not fix) problems discovered in flight test and the related timing of those fixes tends to be almost completely focused on the eventual impact to the production configuration. While that aspect is certainly important it should not be the only criteria. The impacts to ongoing flight test are rarely addressed as aggressively as needed. These decisions tend to be made at high levels for all problems found, regardless of the nature of the problem.

In addition, decision makers with a "process-management" orientation usually grade the system developers based on the overall rate of resolution of problem reports, with only limited (if any) focus on the severity of those problems for either the production design or the flight test impacts. In other words, if the total number of open problem reports was going down, the decision makers were satisfied. The unintended consequence was the developer focus on resolution of lesser problems with easy fixes to show "progress" by more quickly burning down the big list of problems. Fixes to major problems affecting the operational capability and the ability to accomplish productive flight testing were delivered slower than needed, causing significant disruptions to flight test, with related cost increases. The sense of progress portrayed by oversimplified tracking of problem resolution was far less than the actual progress. Another side effect of the narrow focus on burning down the big pile of problem reports has been the lack of focus on the duration between problem identification and problem resolution. Some important problems have been "stuck" awaiting fixes for 1 to 2 years or longer. Although it may have been possible to continue flight testing over that entire duration, much of that testing was not as productive as it should have been because it was accomplished with known deficiencies, requiring significant amounts of inefficient workarounds and unplanned repeats.

DISCUSSION: Flight test, particularly on large development programs, has been especially prone to discovering and documenting many problems. Large programs can have hundreds of open problem reports. The sheer magnitude of problems has choked the typical decision making process for implementing fixes. Part of the problem has been the going-in presumption that "Flight testing is only a verification step to confirm that the system meets requirements… and maybe fix a few unexpected problems". That very common programmatic philosophy ignores the fundamental reality that flight testing on every new aircraft project is a very difficult development effort, not just a simple verification effort. When the entire project is predicated on a highly optimistic assumption, the developers are just not prepared for the large number of problems actually encountered. The unfortunate and repetitive pattern has been that many people on almost every new project seem to truly believe that they have found the magic approach that will prevent them from falling victim to the historical pattern. Any attempt to plan for a more realistic outcome has been viewed as overly pessimistic and inconsistent with program principles. It would be far better to adhere to the time-honored philosophy of "Hope for the best… but plan for the worst".

Once the reality of the escalating number of problems requiring fixes has sunk in, most projects have eventually developed a reasonably aggressive process to provide those fixes. However, it has sometimes taken a year or two after first flight to develop that aggressive process. In the meantime, the impacts on the flight test process have been highly traumatic. The end result was dozens of unplanned software-hardware combinations. In some cases the eventual rate at which those new software-hardware versions were provided was truly astonishing (every few weeks sustained over several years). Although very impressive, that high rate reflects more of a "panic" response than a precognitive plan. The entire process would have been more effective if a smaller but reasonable number of interim updates was planned from the very beginning (see Lesson Learned "CA6: Plan for Adequate Number of Software Revisions")

The consequences of problems discovered in flight test can range from very minor to showstoppers. The very minor problems don't have serious consequences to the operational capabilities, and usually don't delay significant portions of the flight test. Showstoppers could be as serious as grounding the aircraft and/or inhibiting much of the progress of flight test. In between those extremes there is an entire spectrum of problems that have significant impacts on the intended operational capabilities, ongoing flight test progress, or both.

As one example, if a given problem has completely stopped envelope expansion, it gets a lot of attention and leads to an "emergency" update intended to enable resumption of envelope expansion. However, there also tend to be many problems that have more subtle, but still significant, impacts to ongoing flight test. Those "smaller" problems often get lost in the shadow of all the "big" problems. Those "smaller" problems might not shut down envelope expansion completely, but they often preclude continued testing along a given path. In those situations the flight test community attempts to find alternate paths (see Lesson Learned "FF2: Flight Test Flexibility"). However, while testing along those alternate paths, test points get accomplished that will need to be repeated when a fix for the problem eventually shows up. The longer the delay in a fix reaching flight test, the bigger the workload required to repeat test points.

The decisions on if and when to fix problems are typically made at levels higher than needed in many cases, which exacerbates the timely resolution of the problems for flight test. On a typical EMD-type program, a majority of problems with individual subsystems are clear-cut, non-controversial issues with straight-forward fixes. These should be addressed by the people who know that subsystem best (i.e., at the subsystem IPT level), but the "racking and stacking" of fix priorities often occurs at much higher levels. Once problems are elevated to those levels, the decision makers often have excellent operational backgrounds or strong process-management experience, but may not have any flight test background or may have insufficient technical depth to truly understand the full implications of the problems.

Some problems of the straight-forward nature cut across boundaries for two or more subsystems or technical disciplines. These require somewhat higher level coordination of the fixes for the problems to ensure that the fix still meets the needs of all stakeholders. Although problems of this straight-forward nature should be handled at the lowest level practical (via coordination between the affected IPTs), the decisions will inevitably be made at higher levels if elevated prematurely. The goal for any project should be to create an environment that encourages fixing problems at the lowest level practical and to only elevate incrementally as needed by cross-discipline impacts or when the fixes require significant tradeoffs between cost/schedule/capability.

The successful completion of any flight test program relies on the timely resolution and retest of the most important problems. However, resolving so many overlapping problems can be very difficult, so process-management people frequently attempt to simplify the process by focusing on the rate of resolution of these problems (burn-down) and the rate that new problems are added. The developer is incentivized to show rapid burn down of the problem reports. However, in most cases this leads to the developer focusing on lesser problems with easy fixes to show "progress" and quick burn down, while the major problems needing significantly more effort to resolve don't get the immediate attention and resources warranted. In some cases this has actually led to a "bow wave" of serious unresolved problems that resulted in significant delays to the overall progress of the flight test program.

Tracking the burn-down of problem reports without consideration of what has been resolved in terms of the relative contribution toward the operational capability needed and the impacts of the problem on flight test progress has had the additional unintended consequence of premature closure of problem reports. There has been a strong tendency to close problem reports once a fix has been designed and evaluated in a lab or simulation, but before it has been verified in flight test. Normally, if a problem was first discovered in flight test, it has been best to verify the fix in flight test. There may be exceptions where it may be adequate to verify and close a problem report based only on lab or simulation, but the flight test team needs to concur that the lab/simulation setting is appropriate and sufficient for that particular problem.

In extreme cases, middle management has actually directed test engineers in labs to stop writing new problem reports to avoid growth of items being tracked by the metrics used to grade those managers. That is the direct opposite of the desired behavior since fixing problems early using lab testing is the first line of defense against finding those problems in flight test. The programmatic philosophy for testing fixes and the related metrics should be very clearly established to avoid similar bad behaviors.

The descriptions in USAF problem reports (Deficiency Reports [DRs] and Watch Items [WITs]) typically have not adequately addressed the consequences of the problem on flight test progress nor adequately explained the operational impacts of deeply technical problems. The originators of those DRs and WITs have often communicated the nature of the problem in highly technical language unique to a given discipline, and assumed that the reader had a similar depth of technical understanding and would be able to extrapolate to infer the operational and flight test impacts. Unfortunately those assumptions were often invalid, even though that level of understanding was vital as a part of the decision process for when and how each problem was addressed.

Past projects have also relied heavily on "word-of-mouth" communication of flight test impacts versus incorporation into a documented process. As discussed earlier, the end result has been that the decision process has focused too much on the eventual need to fix the problem with respect to perceived operational capability, not the near term flight test impacts. The best way to avoid those undesirable results is for the originator of the flight test DR or WIT to very clearly communicate the impacts for both flight test progress and operational capability. This can be more of a challenge than it might seem because the originators are sometimes narrowly focused young discipline engineers with limited aviation background and no operational experience. Those engineers may require considerable help from their more experienced co-workers and local leaders to properly communicate the impacts.

There has also been a lack of adequate tracking between initial identification of a problem and the final resolution of that problem. Too often the contractor's internal problem report had been closed even though the problem had not been fixed and a new problem report had to be initiated for the same basic problem. Even though the USAF DRs and Watch Items have contained a date for the initial problem identification, there was often no metric to track how long it took for an adequate fix to be incorporated. The lack of a metric to track and report fix duration has been one of the factors that led to durations of 1 to 2 years between discovery of the problem and final resolution. The mere existence of a metric to track fix duration could provide enough management insight to avoid fix durations of 1 to 2 years or longer.

A big reason these unfortunate situations have occurred on past programs has been because the system developers were not adequately focused on efficient progress through flight test. There was no incentive to deliver timely fixes for problems with near-term flight test impacts, as long as the operational requirements were eventually met. The end result was that problems with flight test impacts were addressed only when flight test management (including both contractor and government leaders) elevated each problem to very high programmatic levels. That process was often very contentious since it necessitated "unplanned" updates. This situation could be somewhat alleviated if flight test-oriented updates were part of the plan instead of an exception that required disruption of the previously planned major updates (see Lesson Learned "CA6: Plan for Adequate Number of Software Revisions").

APPROPRIATE ACTION: Each flight test project needs to develop more explicit methods to find a good balance between addressing important problems that need to be fixed for the production configuration while also providing timely fixes to problems with near-term flight test impacts. Flight test management needs to aggressively advocate that fixes for flight test problems be built into the overall plan, not just developed as an afterthought. This should include working with the program office and the contractor to ensure that enough lab capacity and staffing is in place to enable a normal V&V process for multiple flight test-oriented updates every year, not just an occasional update for new capability.

The action that is most likely to help achieve an efficient flight test program is to encourage quick resolution of problems at the lowest level practical (within an approved philosophy and set of guidelines). If every minor problem on every subsystem has to be reviewed and approved at the highest levels of coordination, that is a recipe for shockingly slow progress. The highest levels of coordination should be reserved for the most important or most difficult (i.e., expensive) problem fixes. The majority of problems with individual subsystems that are clear-cut, non-controversial issues with straight-forward fixes should be addressed by the people who know that subsystem best (i.e., at the subsystem IPT level). When straight forward problems cut across boundaries between two or more subsystems or technical disciplines, and require coordination of the fixes across these disciplines to ensure that the fix still meets the needs of all stakeholders, resolution should still be handled at the lowest level practical (via coordination between the affected IPTs). See Supplemental Discussion for additional insights.

The flight test impacts of a given problem need to be clearly identified within each USAF problem report (DR or Watch Item). Examples of the types of categories that might serve this purpose are:

- Primary envelope expansion paths prevented or inhibited
- Secondary envelope expansion paths prevented or inhibited
- Integrated System Evaluations (ISEs) cannot be executed as planned
- Flight test operations rendered less efficient than planned
- Subsystem cannot complete intended flight test points

There are probably other examples that would work better for a given project but the important point is that these flight test impacts be part of the decision process for when and how each problem gets addressed.

In addition, flight test engineers should also be trained to explain problems in terms of operational impacts and not limit their write-up to deeply technical explanations. This may require considerable help from co-workers, local leaders, pilots, Operational Test, contractors, and others. In cases where the operational impacts are especially obtuse, it might be appropriate to explain the technical issues in the context of possible impacts to the Airworthiness Certification process.

Each project should develop some type of usable metric to track and provide visibility for the total duration between initial identification of a problem and the final resolution. Clearly, tracking of that duration needs to be associated with the criticality of the problem. Some minor problems (or proposed enhancements) may not even require a fix during EMD and should not be tracked with the same gusto as problems with big operational or flight test impacts. The main point is to provide adequate data so that decision makers can assess each issue within its context, and assess if the total fix duration is appropriate for the criticality of that particular problem. The mere existence of a metric along these lines may be very helpful to keep problems found in flight test from languishing without fixes for excessive durations. It will be important for those metrics to track the duration from the original identification of the problem without resetting that metric to zero every time an incremental fix is incorporated and fails. It may also be appropriate to track the total number of attempted fixes for a given problem to make it clear when the problem is more difficult than advertised and when the existing fix process just isn't working. The flight test community usually has a pretty good idea when the proposed fix process amounts to little more than "Let's try this and see if it works". The proposed tracking metrics may help make a more solid case than just a bunch of opinions from flight testers.

None of the philosophy advocated above will matter if there is insufficient capacity in terms of personnel and labs for fixing unexpected problems (in addition to providing new capabilities). The Verification and Validation process for those problem fixes must also be capable of handling a higher workload and have an efficient path for blending intermediate versions with planned new versions. Although these things can be aggressively advocated by flight testers, the decisions will remain with others.

SUPPLEMENTAL DISCUSSION: As discussed above, most problems should be handled at the lowest level practical (via coordination within and between the affected IPTs). These problem categories should not require initiation of a USAF DR. A combination of USAF Watch Items and the contractor's internal problem tracking system should be sufficient in most cases. The contractor's internal problem tracking system is an essential step to get action on actually fixing a problem, but should not be used as a substitute for a USAF Watch Item. The USAF Watch Items provide a mechanism to ensure that a particular problem doesn't get "lost" under a large pile of other problems. Watch Items work best when updated to reflect the results of root cause analyses to help understand what needs to be fixed and when. USAF DRs should only be used when the situation warrants the higher levels of coordination. Examples of those situations are:

- The problem would directly prevent meeting top level system requirements, e.g., those defined as Key Performance Parameters (KPP) or Key System Attributes (KSA) in a System Requirements Document.
- The fix will be very difficult, very expensive, or require significant tradeoffs in the overall weapon system.
- It becomes apparent that the contractor is not fully committed to quick resolution of the problem or if the proposed fix will not achieve the necessary results.

If the fix proposed by the contractor appears to be on a timely and appropriate path, the issue can remain at the USAF Watch Item level. The administrative costs of the DR system should be sufficient reason to avoid writing DRs for every minor subsystem problem. In addition, the DR process is well suited to handling relatively small numbers of important issues but is not well suited to handling the combination of hundreds of major and minor issues.

A recent example of a well-balanced DR/Watch Item problem reporting system has been developed at the 411 FLTS. That system has been very well received by the program office, contractor, and the test community. The 411 FLTS approach would provide a worthwhile template for any flight test squadron. Obviously the concepts used by the 411 FLTS would need to be adapted to the type of project (such as EMD, sustainment, etc.). A key factor that has made the 411th process work so well for all stakeholders has been the excellent visibility provided for each of the participants. Engineers and management at the program office, contractor, AFOTEC, and CTF all have the same access and capability to understand the priorities and status of each Watch Item. See Appendix C for a generic adaptation of the 411 FLTS process along with more descriptions of what makes that process work so well.

FF6: PRUDENT FOCUS OF FLIGHT TEST POINTS

<u>LESSON LEARNED</u>: On large scale envelope expansion flight test programs, the edges of the envelopes tend to present the most surprises, but the scope of testing there typically contends equally with the scope of testing in the rest of the envelope. If the developing contractor has employed rigorous discipline in the development of high fidelity models and simulations, this distribution of test focus does not take prudent advantage of the available test scope. It can dilute the scope of testing in the areas of the highest problem potential, thereby possibly missing detection of problems, or inadvertently encountering severe to catastrophic problems. A more prudent approach would be to consciously take larger steps between test points in the middle of the flight envelope, and focus more explicitly with smaller steps near the edges of the flight envelope or in regions where there is less confidence in predictions. This prudent focus of flight test points hasn't always been a problem, so is captured here as a best practice.

<u>PROBLEM</u>: An excessive amount of flight testing tends to be conducted in the heart of the aircraft envelopes where experience has shown there is the least likelihood of encountering unexpected issues. The scope of testing (number of test points, flight hours, flight test months, etc.) is constantly challenged and tightly controlled on many flight test programs. Near the edges of the envelopes has proven to present the most surprises, where finer steps and more emphasis is needed. But the scope of testing available to focus on these areas has been diminished by the scope of testing conducted in the heart of the envelopes.

<u>DISCUSSION</u>: For this discussion, the phrase "edge of the flight envelope" is not intended to be solely in terms of high Mach number or high airspeed. The "edges" of interest can also be angle of attack, angle of sideslip, 100 percent limit load, transonic, or any other variable that stresses the design to near maximum capability. Therefore the critical "edge" can be encountered in the "heart" of the airspeed-altitude envelope for a given aircraft. For example, 100 percent of limit load on a control surface can be encountered in the middle of an altitude-airspeed envelope simply because the flight control system architecture allows the largest control surface deflections and angular accelerations at those conditions. The most important takeaway from this Lesson Learned is to understand how much margin (away from an undesirable result) is available in each region of the flight/maneuvering envelope for a given discipline and to design the test matrices accordingly. This approach may seem like an obvious test planning prerequisite, but this Lesson Learned is in this handbook because the issue has been a recurring theme on multiple test projects.

The classic approach to large scale envelope expansion flight testing has been to start in the "heart" of the flight envelope and step out in defined increments towards the limits of the envelopes. In some cases the step size between test points (in terms of Mach number, airspeed, AOA, load factor, or another variable) has been fairly consistent across the entire flight envelope. In particular, some initial contractor test point matrices have been very symmetric in every dimension, simply because they were forced to draft a test plan on a compressed schedule and didn't have the opportunity to really think through the tradeoffs and priorities. Couple this with the inevitable and repeating pressure on the test program to scale back the scope of the flight testing (typically motivated by an effort to get back on schedule), and the result has been fewer build-up points and reduced concentration of effort closer to the edges of the envelope. Flight testers on more than one project have looked backwards and wished they had spent less time in the heart of the envelope.

Recent flight test history has shown that when the developing contractor has employed rigorous discipline in the development of high fidelity models and simulation, the resulting flight test predictions are typically quite good in the heart of the envelope (for a given variable). The dependability of those predictions has tended to degrade as the edges of the envelopes were approached and system nonlinearities and other factors came into play. Therefore, more deliberate focus is warranted in the areas near and at the edges of the envelopes or where there is less confidence in predications (such as transonic). The classic approach to envelope expansion flight test can be tempered with the use of high fidelity flight test predictions. The amount of testing conducted in the heart of the envelope can be reduced, taking larger steps/increments, as long as the results are matching flight test predictions within predetermined bounds. This frees up flight test scope that can be applied near and at the edges of the envelopes where finer steps and more emphasis is needed. This enables approaching the limits in finer steps, and with more configurations, while making comparisons with point-by-point predictions along the way. The inevitable surprises that are found near and at the boundaries of the envelopes, and with different configurations, are approached slower and more methodically, increasing the likelihood of finding the problem before it reaches the point of being severe or even catastrophic.

The fundamental message in this Lesson Learned should not be interpreted as advocating that extra testing needs to be accomplished at the edges of the flight envelope. Flight testing at the edges tends to be the most expensive, incurs the highest safety risk, and often requires the most difficult test techniques. Therefore the scope of testing at the edges needs to be thoroughly assessed to be sure that the test points identified are both necessary and sufficient. Any flight testing near the edges of the flight envelope that is redundant or of questionable value should be challenged to ensure a valid need. Given the costly and risky nature of testing at the edges, the "challenge and validate" process has tended to be a normal part of most test programs so no special changes are advocated. The fundamental message of this Lesson Learned is intended to focus on an aspect that has tended to be relatively unchallenged on many test programs... more test points than necessary in the heart of the envelope. The fundamental intent is to ensure that the overall test matrix is appropriately balanced, possibly minimizing external pressures to reduce test scope just as the most difficult test points are approached.

APPROPRIATE ACTION: USAF flight testers should work closely with the developing contractor to determine how much rigorous discipline has been applied to development of high fidelity models and simulations. Many contractors have applied a very high level of rigor to creation of representative models and simulations during early development and throughout flight test. Other contractors have applied an enormous amount of initial effort developing hi-fidelity models based on a combination of wind tunnel testing and Computational Fluid Dynamics, but have fallen short when it comes to updating those models based upon flight test results (see Lessons Learned "CA9: Update System Models with Flight Test Data").

Once an understanding of model fidelity exists, prudent choices of where to concentrate flight test points should be made. Testing where predictions are typically good should be scaled back, but using cross checks that the system is behaving as predicted. Predetermined criteria need to be coordinated to deal with unexpected results (see examples in Lessons Learned FF8: "Residual Oscillations", and FF9: "Rate Limited Control Surface Oscillations" for more detail and examples along with the concept of Flight Test Continuation Criteria in FF4 "Multi Discipline Envelope Expansion"). Finding the right balance can be tricky. When a brand new aircraft design is being tested, there will be a tendency to define too many test points in the heart of the envelope, at least until confidence in predictions has evolved. Testing where predictions typically break down (near and at the edges of the envelopes) should be reinforced with a robust concentration of test points involving build-up in both the test conditions and the test configurations (e.g., different store loadings, different control system states, etc.). This concept applies anywhere in the envelope where predicted performance is suspect.

Some flight test projects have followed this best practice but many others have not. A common telltale indication that flight testing for a given discipline has been developed without adequate attention to the tradeoffs is when the test matrix is nice and symmetric across the flight envelope with relatively even spacing between test points. When that characteristic is observed, it should serve as a "flashing beacon" that more attention is warranted. In that situation the test matrix should be actively challenged within test working groups. Based on the trends typical for a given discipline, key questions should be asked to obtain justification for densely packed test points. For example: "The predicted data is very linear across this entire region… why are so many test points needed? Alternatively: "The predicted aerodynamics (or gain schedules, etc.) are very non-linear across this other region… why aren't there more test points in that region"?

FF7: QUALITATIVE EVALUATIONS

LESSON LEARNED: Flight test evaluations that are dependent on pilot techniques and opinions need to be treated differently than quantitative evaluations that can be resolved by inspection of a few key parameters. The most common types of qualitative evaluations have been accomplished to assess aircraft handling qualities. In general, qualitative evaluations require a reasonable number of repeat evaluations by different pilots in order to minimize the possibility that undiscovered problems might be passed on to the user. In addition, qualitative evaluations using multiple pilots are necessary to ensure that the resulting assessment is not overly influenced by a single dominant pilot due to individual habits / preferences / experiences that drive pilot perspectives (see the example in the SUPPLEMENTAL DISCUSSION section).

PROBLEM: The trend on many recent large scale flight test programs has been towards cutting qualitative flight test evaluations down to a bare minimum set of "spot checks" based on only one or two pilot assessments for a given task/configuration/flight condition. Part of the rationale for those cuts has been that the flight tests are only needed to "verify the model." Towards the end of many development projects, the programmatic pressure to reduce the number of pilots used for qualitative evaluations has become especially strong. This has occurred in spite of the fact that the remaining test points were often the most critical and challenging for that weapon system. If this trend continues, it will become more likely that significant issues will not be discovered until well into deployment or sustainment when fixes will be much more expensive to implement.

DISCUSSION: Model verification flight testing makes sense for quantitative tests but not for qualitative tests. Quantitative flight test results can be compared directly with predictive models. Even if flight test conditions were not identical to the original predictions, the model can be re-run at the flight test conditions to provide a more direct numerical comparison.

Qualitative flight test evaluations do not provide realistic opportunities for similar direct comparisons with predictive models. Ground-based piloted simulators have functioned as the primary predictive model for qualitative evaluations and have been used very effectively to develop new aircraft flight control systems prior to flight test. However, any ground-based piloted simulation has known limitations. Even the most expensive hardware-in-the-loop simulations with hi-fidelity models and hi-resolution visual display systems do not necessarily translate results into the real-world flight environment. Motion-based simulators may add a certain amount of realism in some situations but the results have been debatable and have often been viewed as not worth the additional cost and complexity. Ground-based piloted simulations will always need to be supplemented by an adequate amount of operationally representative in-flight evaluations. Qualitative in-flight tests are valid and necessary because quantitative predictions may or may not manifest characteristics that are objectionable to pilots.

For many decades, developmental flight testing has been viewed as a rigorous process that thoroughly evaluated a broad spectrum of configurations and flight conditions to cover all aspects that would be needed by future operators. The goal had been to fix problems as early as possible while that process was still relatively inexpensive, and provide an end-product that had a high probability of meeting user needs. The escalating cost of flight testing has resulted in drastic cuts to qualitative evaluations without any serious study of the tradeoffs. Typically, someone in a position of authority has simply accepted the risk of those cuts in order to obtain short-term cost savings without any real understanding of the long term implications. No formal study has been accomplished to assess the relationship between reductions in qualitative evaluations and the number of issues that have escaped detection during developmental testing and have required fixes as a result of operational testing or end-user complaints. The fundamental question that will need to be addressed is: "How do decision makers know when qualitative evaluations have been cut too much?"

It is acknowledged that the rigorous qualitative evaluation process that has been implemented on historical flight test projects has become unaffordable. However, the traditional implementation of "three or more pilot evaluations" for each combination of task/configuration/flight condition is rapidly being replaced with the philosophy that "an evaluation by a single test pilot is good enough." On at least one recent major EMD program, the pressure to reduce flight test points became so great that an Executive Independent Review Team was assembled with the goal of justifying that all qualitative test points would be considered complete after a single test pilot's evaluation. However, the independent team confirmed that the original scope of testing was appropriate and that multiple pilot qualitative evaluations were still necessary. Since qualitative evaluations are very subjective, the overall result is heavily influenced by statistical relevance. The fewer the number of evaluators, the less statistically relevant the results. There needs to be a practical balance between the expense of achieving very high statistical relevance and no statistical relevance.

The trend to reduce qualitative evaluations is expected to continue as long as programs can provide the appearance of avoiding major undetected issues. If anything, this trend has been building momentum as promoters on new projects have used the wording in recent flight test plans from similar projects to justify reductions in the number of pilots planned for qualitative evaluations on the new project. Those promoters have only been interested in cutting the scope of the program, and were eager to accept overly optimistic expectations that no problems would be missed. It probably won't matter that many minor undetected issues will be more expensive to fix because no one will be tracking that relationship. The only thing that might reverse the trend will be an obvious connection between a major undetected issue and reductions in the number of pilots providing qualitative evaluations for a given task/configuration/condition.

APPROPRIATE ACTION: Flight test planners on future projects will need to work with contractor and program office counterparts at the discipline level to establish a fundamental philosophy towards qualitative flight test evaluations. Assuming those counterparts have had practical experience with qualitative evaluations they will undoubtedly understand the need for more than a single pilot evaluation. A united front of experienced designers, program office engineers, and flight testers will always be the best defense against uninformed directives to reduce pilot evaluations to inappropriate levels.

One approach would be to start with the traditional "three or more pilot evaluations for each combination of task/configuration/flight condition." That approach has a lot of successful history behind it. One study by a major airframe contractor concluded that 6 pilots would provide the right sample size for obtaining pilot evaluations. Starting with three pilots as the baseline would allow a little room to respond to the inevitable imperative to cut flight test points. However, it is expected that on some future projects, the requirement for evaluations from three pilots for every qualitative evaluation would not even survive the initial approval process for a draft test plan.

If the best efforts of the cohesive discipline team remain inadequate to satisfy test point reduction directives, there are other options that might help salvage a flight test effort. A compromise that has worked reasonably well has been to require a minimum of two pilots as the baseline for qualitative evaluations. However, that baseline has come with the caveat that both pilots have to agree that no improvements are warranted to consider the evaluation of that test point to be complete. If the two pilots disagree about the need for improvements, a third pilot (or more) would need to be called in to help determine the best course of action. Although that approach might be considered acceptable on some projects, there would still be a significant possibility that undetected issues could get delivered to the end-user. The major decision makers would need to be informed of the risk and still be willing to accept that risk tradeoff.

Another possibility is to identify a small number of critical combinations of task/configuration/flight condition that require a more extensive evaluation. Those test points would be the most directly relevant to operational requirements and would require an absolute minimum of three pilot evaluators to reflect that importance. The remaining qualitative test points would presumably be at less critical conditions and therefore could be considered complete after a less rigorous two pilot evaluation, or maybe even after a "spot check" with a single pilot (especially for "build-up" test conditions or conditions with less operational relevance). The main concept behind this philosophy is that every qualitative test point doesn't necessarily require the same number of pilot evaluators.

An important related concept is to consciously build test plans with qualitative evaluations that are aimed directly at the operational uses for that aircraft, in addition to the rigorous handling qualities evaluations. A typical Cooper-Harper rating depends heavily on the Desired and Adequate performance criteria (assuming the pilot uses that rating scale as intended). Those performance criteria are typically selected long before flight testing begins. However, the additional rigor required to obtain a valid Cooper-Harper rating, can also obfuscate characteristics of high operational relevance. Therefore it can be shockingly important to allow developmental test pilots to "put the aircraft through its paces" without being tied to inflexible procedures. This type of unscripted testing can uncover significant issues. Although this type of unscripted testing has often been viewed as a "graduation exercise", care should be taken to avoid the perception that this type of evaluation must wait until "near the end" of a development project. In fact, the sooner this type of testing is accomplished the better. There is no need to wait until the flight envelope has been completely cleared in every dimension. Early evaluations of this type will provide the designers with more time to respond to any unexpected discoveries.

SUPPLEMENTAL DISCUSSION: Paradoxically, at the same time that qualitative evaluations have been arbitrarily cut to the bare minimum or below, repetitive quantitative evaluations have been expanding with the goal of obtaining statistically defensible test results. This has occurred in spite of the fact that quantitative flight tests tend to produce essentially the same results when repeated at the same set of test conditions. Small variations in those test conditions can result in unsurprising and usually predictable variations in those results. Nevertheless, the overriding goal of achieving statistically defensible results has trended towards more complex test matrices in an attempt to achieve a higher confidence level in test results that could vary due to changes in test conditions.

There are very few examples where the same rigorous statistical methods have been applied to qualitative evaluations in order to identify the proper number of repeat evaluations that would actually be necessary to have a sufficient level of confidence in the final product. If that type of analysis were to be conducted, it is likely that the number of required pilot evaluations would be greater than one, and it would not be surprising if the statistically correct answer was greater than three.

Example: During development of the power approach control laws for a new fighter, the primary test pilot supporting the development in the simulator had significant influence on the design. He was a former Navy test pilot whose piloting techniques were very strongly influenced by the Navy techniques for landing aircraft on an aircraft carrier. By the time other test pilots with Air Force background had the opportunity to assess the PA characteristics of the fighter there were significant complaints about the PA handling qualities when typical land-based landing techniques were employed. Very expensive and time consuming rework of the PA control laws and side-stick controller characteristics were required. A larger sample size of pilots with a spread of experience level and experience base involved in the development would have prevented a long delay in addressing this issue.

FF8: RESIDUAL OSCILLATIONS

LESSON LEARNED: On multiple EMD projects over the last few decades, residual oscillations have been one of the most common types of Flight Controls/Flying Qualities problem to be first discovered in flight testing. Each future test team needs to be prepared for the likelihood of discovering similar residual oscillations and have plans in place for responding to these potentially disruptive events.

PROBLEM: Flight test teams have not always been properly prepared when residual oscillations were discovered, usually during envelope expansion testing. When these residual oscillations were first discovered, they tended to have small amplitudes and were not especially objectionable to the test pilot. However, the fact that the system oscillates when not predicted is usually a sign of a basic instability. In some cases, envelope expansion was allowed to continue with no further difficulties and a design fix was eventually incorporated. In other cases, envelope expansion was allowed to continue but the test aircraft encountered some other mechanism (turbulence, jet wake, higher control system gains, nonlinearities at new flight conditions, etc.) and aircraft loads limits were exceeded, sometimes with highly disruptive programmatic impacts.

DISCUSSION: Although the normal design process for Flying Qualities and Flight Control Systems has improved dramatically over the last few decades, residual oscillations have still been a recurring problem. The normal design process has been reasonably successful at filtering out many historically common Flying Qualities issues, but residual oscillations have had a tendency to escape detection during analyses, simulations, or lab testing and have only been discovered during flight test. This occurs because the aircraft models in simulations do not always include the most up to date aerodynamic models, mechanical nonlinearities (such as in simplified models of the actuators), or other real-world characteristics.

For this discussion, residual oscillations include continuous aircraft oscillations in any axis that are not sustained by pilot inputs. Therefore, Pilot-in-the-Loop Oscillations (PIOs) are not included. A residual oscillation may have been triggered by pilot action (such as a rapid control input with significant amplitude) but continues even when pilot inputs have returned to zero. Residual oscillations often occur on aircraft with "fly-by-wire" flight control systems (either digital or analog) that have high gain feedback loops. However, residual oscillations have also occurred on reversible flight control systems with mechanical linkages.

The types of residual oscillation included in this discussion are: Limit Cycle Oscillations (LCOs), Structural Resonance (also known as Structural Mode Interaction (SMI), Ground Resonance Testing, and other names), and typical neutrally damped (or very lightly damped) oscillations related to insufficient gain or phase margin. Oscillations with control surface rate limiting do not normally fit well into the category of residual oscillations because they tend to be more transitory and much larger in amplitude and are therefore not included. That type of rate limited oscillation has resulted in some of the most insidious and disruptive flight test incidents and is discussed in a separate Lesson Learned (see "FF9: Rate Limited Control Surface Oscillations"). Oscillations that are purely due to structural modes (without control system interaction) are not included. Aircraft flutter and Aero-Servo-Elastic (ASE) oscillations are not included.

Oscillations due solely to insufficient gain or phase margin are included in this Lesson Learned but are not the primary focus. On most recent development projects, oscillations of this type have been minimized during the normal design process and have rarely resulted in significant problems during flight test. When oscillations of this type have been discovered during flight test they tend to be lightly damped (i.e., the amplitude reduces over time after an initial triggering mechanism), not neutrally damped where the oscillation continues at about the same amplitude until conditions change. On some projects oscillations of this type have occurred in flight test due to mispredicted aerodynamics but the fix tends to be relatively straightforward… update the aerodynamic model and redesign the flight control system as needed.

On relatively recent projects (within the last 2 to 3 decades), the more common types of oscillations that have slipped past the normal design process and have first been discovered in flight test are LCOs and structural resonance as described in the Supplemental Discussion below. Those oscillation types tend to be harder to predict during the normal design process because models of aircraft components do not always reflect actual in-flight characteristics.

The first “line-of-defense” against the occurrence of these residual oscillations is adequate ground testing on a test aircraft or in an iron bird simulation facility as described in the Supplemental Discussion below. When the ground test first line-of-defense is insufficient and these residual oscillations are still discovered in flight test, a well prepared test team is the second line-of-defense and needs to be focused on minimizing the potentially negative impacts. That aspect is the main focus of this Lesson Learned as described below and in the Supplemental Discussion.

APPROPRIATE ACTION: Most airframe contractors do an excellent job of planning and executing structural resonance ground tests. The effectiveness of planning and execution for LCO ground testing has been more sporadic. During the ground test period prior to first flight the main role of the flight test team is to work with the contractor and technical representatives from the program office to ensure adequate coverage of test conditions, verify that test procedures will accomplish the intended objectives, participate in representative ground tests, and understand how test results might impact upcoming flight tests. After these ground tests are accomplished it normally takes at least one design iteration to adequately adjust structural filters and flight control system gains to minimize the potential for structural resonance and LCOs. Even after the new filters and gains have been implemented, the flight test team still needs to understand which conditions within the flight envelope result in the lowest stability margins.

Structural resonance and LCO ground test results will also be used by USAF Airworthiness authorities as a portion of the formal assessment to provide an Airworthiness certification to begin flight testing. This certification establishes a good starting point for the flight test program but residual oscillations could still occur during envelope expansion. Key members of the flight test team need to be prepared for the occurrence of residual oscillations and consistent procedures need to be captured within project documentation along with team-approved actions after oscillations occur.

Residual Oscillation procedures within project documentation need to be established in a way that finds the right balance between consistent team reactions regardless of which individuals are in the control room, and preservation of sound engineering judgment. When that balance has not been found, past projects have been impacted by both extremes. Some projects have relied too much on engineering judgment which was excessively influenced by high pressures to continue envelope expansion and led to highly disruptive mishaps. Some projects have defined inappropriately tight and arbitrary criteria that required unnecessary test termination and the corresponding emergency coordination of new procedures to allow envelope expansion to resume.

Many projects have used pre-established criteria for residual oscillation magnitude and/or number of cycles with good results. These criteria have worked best when based upon the specific aircraft and its control system in terms of whether it can safely check the oscillations when the forcing condition is removed. For example, an aircraft with highly negative static margin will be at more risk of an oscillation quickly growing to dangerous levels than would an aircraft that has positive static margin. It has become common for many types of modern aircraft (fighters, bombers, and transports) to be designed with neutral to slightly negative static stability. This makes those aircraft more susceptible to residual oscillations than historical aircraft with positive static stability. It has been impressive how well test teams have defined logical and useable residual oscillation criteria AFTER a mishap or near-mishap. It is hoped that this discussion will inspire more test teams to define reasonable criteria BEFORE those situations are encountered.

The simplest test team procedure is to cease envelope expansion at the first sign of a lightly damped oscillation. That type of procedure provides for a somewhat unequivocal test team response and is most likely to avoid continuing into a mishap situation. However, for some aircraft that type of criteria may be too conservative and the criteria for “lightly damped” would still need to be well defined to avoid ceasing envelope expansion prematurely.

Once the mechanism for a residual oscillation has been evaluated and understood, the test and design teams may have high confidence that the oscillation is indeed bounded and will not degrade at other conditions. In that case, a decision will still need to be made to attempt a fix... or not. Some of the criteria listed below may help guide that decision. If there is no reason to fix a given oscillation to meet a specification or other requirement, pilot opinion regarding mission impact should ultimately be the deciding factor.

More specific examples of residual oscillation occurrences and criteria are provided in the Supplemental Discussion Section below.

SUPPLEMENTAL DISCUSSION: In-flight occurrences of structural resonance tend to be relatively unambiguous since these situations are usually very noticeable to the pilot and the test team response is usually unanimous… cease envelope expansion. Pilot qualitative descriptions of structural resonance occurrences have included terms such as “rough ride” and “gravel road”. Structural resonance frequencies tend to be relatively high (i.e. 2 to 5 Hz) and the distinction between oscillations that are sustained versus lightly damped is essentially irrelevant. Structural resonance has typically been discovered in flight test due to a combination of unexpected interaction from aircraft structural modes through flight control system sensors (such as accelerometers or rate gyros) but aircraft aerodynamics are not normally involved. Although it is theoretically possible to find structural resonance conditions that occur within an “island” (where all the surrounding conditions do not exhibit resonance), it is more typical to find cases where structural resonance gets worse in some dimension (e.g., greater amplitude with increasing airspeed/dynamic pressure, increasing load factor/angle of attack, etc.). Depending on the specific circumstances, some ambiguity may remain and a predefined criterion can be important to help provide an important “outer bound” to engineering judgment. Additional background on structural resonance is provided below the following LCO discussion.

In-flight occurrences of LCO can be much more ambiguous. The amplitudes may be so small as to be essentially unnoticeable to the pilot. When LCOs are noticeable to the test pilot, qualitative descriptions have included terms such as “pitch gallop” or other innocuous sounding terms. The first indication is typically observed by an engineer in the control room observing the tell-tale oscillation on an amplified scale for a control surface position or aircraft response parameter such as load factor or pitch attitude (if the scale of the time history plot in the control room is inappropriate, an engineer might miss the LCO entirely… until post-flight data review or until the amplitude becomes larger at a different flight condition). The test team response is often not unanimous. The engineer that spotted the LCO often wants to cease envelope expansion until the mechanism is better understood. The test pilot and test conductor typically want to continue envelope expansion since they can’t see anything wrong. It is not uncommon to have varying opinions within the Flying Qualities discipline, even between very experienced engineers and even within the contractor engineers in the control room. This is exactly the situation where a predefined LCO continuation criterion is necessary. Many previous projects used continuation criteria based upon the observed damping after an excitation. However, since a typical LCO has no damping at all, criteria based only on damping would preclude further testing along that path for ANY LCO. That may be the intent of the test team for a given project. However, it may still be prudent to continue flight testing for small amplitude LCOs… but how small is considered acceptable? The following paragraphs provide some philosophy for consideration.

Useful LCO continuation criteria can be derived from existing documentation on most projects. Almost every USAF aircraft development project has defined a level of residual oscillation that is considered acceptable in the final design. The background info for selecting values for acceptable residual oscillation levels was defined in older flight control system specifications such as, MIL-F-87242 (*Military Specification: Flight Control System General Specification For*) and MIL-F-9490D (*Military Specification: Flight Control Systems – Design Installation and Test of Piloted Aircraft General Specification For*) and the more recent Joint Service Specification Guide for Vehicle Management Systems (JSSG-2008A) (*Department of Defense Joint Service Specification Guide: Vehicle Control and Management System*) (references 6, 7, and 8). Values for acceptable residual oscillations that have been tailored for a particular aircraft and associated mission tasks should be documented in whichever flight control system specification applies to that specific project. Values that have been previously required in JSSG-2008A were:

- 0.04 g's vertical acceleration (peak-to-peak)
- 0.02 g's lateral acceleration (peak-to-peak)
- 0.3 degrees roll and yaw attitude (peak-to-peak)
- Pitch attitude values were specified to match other flying qualities specification criteria

If the required residual oscillation values for a given aircraft are truly considered acceptable in the final product, it would be reasonable to conclude that those same values would be acceptable to continue flight test envelope expansion… as long as there is a basic understanding of the mechanism and the test aircraft configuration is not inherently unstable. Aircraft configurations with significantly negative static margin will be more susceptible to situations where a small residual oscillation could rapidly increase in magnitude over a small range of flight conditions. It would not be reasonable to continue flight test envelope expansion if the residual oscillation values were significantly greater than the required values for the final product.

One approach that has been successfully used on several new aircraft development projects has been to allow flight test envelope expansion to continue if limit cycle residual oscillation values were no greater than twice that required for the final product. After unpredicted in-flight limit cycle occurrences on various projects, the members of the Flying Qualities team were nearly unanimous on judging the residual oscillations relative to that "double the spec amplitude" criterion. If the oscillation magnitudes were less than that criterion, most team members tended to agree that it was safe to continue. This enabled gathering of essential data necessary to fix the problem. If the oscillation magnitudes were greater than those criteria, most team members agreed that continued envelope expansion was unwarranted given the higher risk of proceeding. That approach has worked because the criterion was discussed in detail and coordinated with all stakeholders ahead of time and not invented in the control room. The "double the spec" criterion provides an excellent place to initiate that discussion within a discipline test team.

The LCOs are typically bounded in amplitude by some kind of non-linearity. Some residual oscillations may appear to be bounded but may only be limited in amplitude because the "bad" combination of flight condition and flight control system gain has not yet been encountered. Even if a particular oscillation has been confirmed to be bounded at the conditions tested (because the non-linear mechanism is well understood) the oscillation may still grow in amplitude at other conditions. These LCOs have typically been discovered in flight test due to a combination of mispredicted aerodynamics and un-modeled flight control system characteristics but normally have not involved structural modes. These LCOs tend to occur at relatively low frequencies (roughly 1 to 2 Hz).

An excellent source for structural resonance and LCO theory and test techniques can be found in AFFTC-TR-76-15, *Flight Test Development and Evaluation of a Multimode Digital Flight Control System Implemented in an A-7D (DIGITAC) Volume 1* (reference 9). Back when that A-7D DIGITAC report was written in 1976, it was common to minimize in-flight occurrences of LCO by accomplishing limit cycle ground testing on a representative test aircraft. When limit cycle oscillations were discovered during those ground tests (quite common since designs tended to migrate towards high gain flight control systems to meet system requirements) it was fairly straightforward to develop a redesign to avoid the oscillations. However, since test aircraft are highly valuable assets and on-aircraft ground testing tends to cover many aspects on a very tight schedule, most limit-cycle testing has migrated into “Iron Bird” simulation facilities. Either method can provide an acceptable test venue but each project needs to make a thorough technical assessment of the tradeoffs.

Iron Bird simulation facilities can provide an acceptable venue if the proper flight control computer hardware is in the loop, has the planned flight test software version installed, and if the control surface actuation system (including any hydraulic components) are good representations of the test aircraft. Contractors typically go through extraordinary efforts to provide hi-fidelity iron bird facilities so they are often acceptable venues for limit cycle testing. On-aircraft ground testing will always provide the most representative test venues for limit cycle testing but the additional cost and schedule impacts may not be worth the higher fidelity. It may be worth consideration to supplement a thorough limit cycle test using an iron bird with some well-targeted limit cycle “spot checks” on a test aircraft. If those on-aircraft spot checks correlate well with Iron Bird tests, that result would help improve confidence that important characteristics were not missed on the iron bird.

Regardless of the limit cycle test venue (Iron Bird or test aircraft) it will be important to close the aerodynamic feedback loops around all of the actual hardware by using a reasonably high fidelity aerodynamic model and accurate equations of motion. In the past, limited computing power necessitated that those aerodynamic models be “linearized” around individual flight conditions for each limit cycle test point along with using simplified equations of motion. Faster computers should enable use of more complex non-linear aerodynamic models and less simplified equations of motion. However there may still be reasons to avoid using the “full-up” aerodynamic models and equations used by other six-degree-of-freedom simulations for that project. Each design and test team will need to determine the right balance and the appropriate fidelity for limit cycle testing.

For whatever reason, contractor organizations seem to have “intermittent understanding” about the requirements and techniques for limit cycle testing. Some contractors are fully onboard with the concept and need no additional prodding to do an excellent job. Some contractors don’t seem to understand that limit cycle testing at some level is essential and view the requirement as “out-of-scope”, probably because the requirement for limit cycle testing can be relatively fuzzy in design guidance. Each test team will need to assess their design team counterparts to determine their level of understanding relative to limit cycle testing. That assessment should guide test team efforts regarding the level of “help” that needs to be applied. As always, coordination with the USAF customer should be used if contractor design teams do not respond appropriately. If the USAF customer also doesn’t respond appropriately, this topic is important enough that it may be necessary to have direct discussions with the relevant USAF Airworthiness technical discipline authority. If the project’s Test Squadron organization is unwilling to support that direct contact it may be necessary to arrange that contact through the engineering home office. Yes, this topic is important enough that exercising these alternate technical coordination paths may be required.

In contrast with limit cycle testing, realistic structural resonance testing can ONLY be accomplished during ground testing on a structurally representative aircraft. Iron bird facilities have no ability to replicate actual aircraft structural modes. Ground Vibration Testing (GVT) on a test aircraft has been an excellent tool (and is a required step) for updating structural models to better represent the actual aircraft structural mode shapes and frequencies. However, even the most highly advanced and updated structural model will likely miss the actual local mode shapes and frequencies as sensed by each flight control sensor. Therefore ground testing definitely needs to take place on a test aircraft prior to first flight in order to ensure flight safety. Substantial portions of that ground testing may also need to be repeated on a more operationally representative test aircraft if significant structural changes occur.

The most important structural resonance testing does not need to include closing loops around feedback sensors using aerodynamic models or equations of motion. Some projects have accomplished on-aircraft ground testing that has essentially been a combination of limit cycle and structural resonance testing. That special type of ground test may be very enlightening and may provide important insights into how those combined loops may interact, but also has the potential to be misleading because of the way that simulated aerodynamic feedbacks to individual sensors need to be combined with the local structural modes as measured by each sensor.

Structural resonance testing does not need to replicate every possible in-flight condition. The main conditions of interest are those where the flight control gains are highest in each feedback path. Finding the highest gain (or smallest margins) in each feedback path is usually fairly straightforward (i.e., inspection of the control laws and gain schedules). However, it can be much more problematic to find the condition(s) with the lowest gain or phase margin while representing the combination of multiple feedback loops all closed at the same time.

Almost all contractor organizations have an excellent understanding regarding the requirements and techniques for structural resonance testing. Contractor design teams normally do an excellent job of accomplishing on-aircraft ground testing and then designing an appropriate structural filter for each sensor path. Once any re-designed structural filters are implemented in the control laws, those contractors are normally fully invested with the need to re-accomplish key portions of the ground tests with the new filters to verify that they actually achieve the desired result. The end result of that normally thorough process has been that relatively few structural resonance conditions have been discovered in flight test. The few exceptions have usually occurred when the matrix of ground test conditions did not adequately cover the spectrum of likely flight test conditions. This is where the experience of the flight test team can help guide the contractor during the development of the structural resonance ground test plan. Ideally that flight test team will include both government and contractor personnel and also include representatives from the Structures disciplines in addition to the Flying Qualities/Flights Controls disciplines.

Another situation where structural mode interactions have occurred during flight test (even after ground testing) has been when the structural first bending mode is not that far from the short period mode. This situation normally occurs on larger aircraft and makes for a difficult design challenge to filter out structural modes without interfering with essential short period flying qualities characteristics. One technique that has been used in that situation is “phase stabilization” instead of “gain stabilization”. This might be a case that suggests additional ground testing on the test aircraft using a combination of simultaneous LCO and structural resonance test techniques. Yet another test technique might be to expand LCO testing on the iron bird to include high fidelity structural models (ideally based on GVT results). The main message is to expect more design effort and more test iterations on projects with similar situations.

An important setup consideration for on-aircraft structural resonance ground testing is the boundary condition for the landing gear. The landing gear or tires are normally replaced with some kind of “soft support system” for any ground test conditions aimed at replicating the structural feedback loops for in-flight conditions. On some aircraft this can be accomplished merely by reducing the tire pressure. On other aircraft a more complicated “air bag” system may need to be implemented. The goal is to reduce the dynamics associated with the landing gear to a lower frequency that does not interfere with test results. Some kind of soft support system is essential for GVT, but may also be important for on-aircraft structural resonance ground tests. However, it seems fairly common for contractor design and test teams to miss the requirement to also accomplish on-aircraft structural resonance ground tests while on the landing gear with tires inflated to normal pressures. The test matrix for this configuration can be much smaller since it only needs to encompass the flight control gains associated with taxi and takeoff (not up and away in-flight conditions).

When a test team first encounters an unexpected oscillation of any kind during flight test, the initial reaction should be to avoid a degrading situation that could lead to a mishap. One of the most common methods for stopping an ongoing aircraft oscillation is to return to a previously cleared portion of the flight envelope (e.g., reduce airspeed, or AOA, or load factor, etc.). That reaction is often documented as part of the safety planning, usually as a General Minimizing Consideration or within a Test Hazard Analysis. That reaction can be a very effective method for dealing with the immediate issue at hand, but does not address the programmatic imperative to continue envelope expansion.

The very first question that a control room test engineer will be asked after getting an unexpected oscillation under control will be something like, “Can we continue to the next test point?” It has always been best if the test engineer was able to answer that question based on a pre-coordinated continuation criterion as opposed to subjective “engineering judgment” developed on the spur of the moment. However, the members (both government and contractor) of many test teams are notoriously reluctant to define a pre-coordinated continuation criterion because “they don’t want their hands to be tied unnecessarily based on an arbitrary criterion.” They are correct but on numerous projects it has been amazing to watch how quickly those same reluctant team members came up with a useful continuation criterion AFTER an in-flight incident had occurred. A useful continuation criterion will not be arbitrary if it is well thought out and coordinated ahead of time.

The phrase “continuation criterion” has been intentionally used instead of “termination criterion.” Both phrases amount to the same thing but have different psychological impacts. The phrase “continuation criterion” implies that if you meet this criterion you can continue. The phrase “termination criterion” sounds more negative because it implies that if you’re just a bit outside the criterion you have to stop.

Some groups of very experienced Flying Qualities engineers have stated that they would stop envelope expansion at the first appearance of any residual oscillation. By contrast, different groups of very experienced Flying Qualities engineers have wanted the option to proceed if they judged the oscillation to be acceptably small (without being constrained by any pre-defined criterion).

Both points of view have merit. The group that stated they would stop envelope expansion at the first appearance of an oscillation has pointed out that the oscillation was not predicted and therefore indicated a problem that needed to be better understood before proceeding with envelope expansion. The group that wanted the option to proceed has pointed out that small oscillations were not a safety concern and that more data would be needed to better understand the scope of the problem in order to design an effective fix. It has been common for engineers with those opposing points of view to be on the same project. The recurring message in this Lesson Learned is for each project to resolve as many of those philosophical disconnects as practical before the occurrence and not rely solely on the vastly different judgment of whichever individual happens to be in the control room on a particular mission. Every test team will need to come up with an effective approach for their project.

One of the trickiest decisions any test team will need to make is: how much LCO and structural resonance testing needs to be repeated after significant flight controls or structural design changes. It is pretty much guaranteed that management from every organization and at every level will be “spring-loaded” to NOT repeat these tests. These tests take a lot of preparation time, a minimum of several days out of the flight test schedule, and require highly skilled personnel and special ground test equipment. The original tests were typically accomplished prior to first flight and significant flight controls and/or structural changes can still be occurring years later. When this happens the highly skilled specialists may have moved on to other projects, and the special ground test equipment may have degraded, become obsolete or unsupportable because the vendor has gone out of business.

Simulations and analyses can be used to accomplish some amount of forecasting, but the fidelity of the relevant models for this purpose will always be in question (as already discussed) (see Lesson Learned "CA9: Update Air Vehicle System Models with Flight Test Data"). As with many flight test issues, it can be very helpful to define a set of criteria ahead of time that can be used to help guide decisions about the need to repeat LCO or structural resonance testing. For example, if the predicted models indicate a change of greater than X dB gain margin or Y deg phase margin, that result may help justify the need for repeat testing. It may be very difficult to get all parties to agree to a set of criteria ahead of time. However, it will probably be worth the effort to help avoid very emotional and heated debates when the question eventually arises. It would also behoove the test team to ensure that adequately skilled personnel remain available and that the special ground test equipment remains capable of supporting repeat testing.

FF9: RATE LIMITED CONTROL SURFACE OSCILLATIONS

LESSON LEARNED: Rate limited control surface oscillations have been a major contributing factor in some of the most insidious and disruptive flight test incidents that have occurred over the last three decades. Each future test team needs to be prepared for the likelihood of discovering similar oscillations and have plans in place for responding to these potentially disruptive, if not catastrophic, events.

PROBLEM: Flight test teams and design teams have not always been properly synchronized or fully prepared with a coordinated response when unexpected rate limited control surface oscillations were discovered, usually during envelope expansion testing. When unexpected oscillations of this type were first discovered the implications were sometimes minimized, usually because the oscillations "didn't seem that much worse than predictions." However, the fact that the system oscillates on the rate limit more than predicted is usually a sign of an inherent stability problem that needs to be taken very seriously. In some cases, envelope expansion was ceased at the first sign of unexpected rate limiting and a design fix was eventually incorporated that enabled envelope expansion to continue. In other cases, envelope expansion was allowed to continue but the test aircraft encountered some other mechanism (turbulence, jet wake, higher control system gains, non-linearity at new flight conditions, etc.) and aircraft loads limits were exceeded, sometimes with highly disruptive programmatic impacts. Rate limited oscillations may not be harmful in some situations, but could lead to disastrous results in others (e.g. ok at nominal CG, very bad at aft CG).

DISCUSSION: Rate limited control surfaces are not a new phenomenon. In fact, control surface rate limits have been part of every aircraft design since the Wright Brothers. On aircraft with classic reversible flight control systems, the rate limits were a natural result of mechanical linkages and the pilot's ability to apply control stick inputs rapidly given the inherent loads, friction, and damping within the system. On aircraft with hydraulic control surface actuators (with or without a flight control computer), the rate limits were typically a physical result of the specific valve design within the actuators and the control surface linkage mechanism. On more recent aircraft the rate limits were often programmed into a flight control computer (either analog or digital) as a way to avoid running into the physical actuator rate limits. That implementation is often intended to reduce wear and tear and increase fatigue life of the actuators.

The most negative aspects of the rate limiting phenomenon didn't really become an insidious problem until those rate limits were included within some type of augmented flight control system such as a Control Augmentation System (CAS) or Stability Augmentation System (SAS). Those negative aspects have manifested themselves most seriously when the gains in the command paths or stability feedback loops have been driven to higher levels in an attempt to achieve a design goal of one type or another.

Control surface rate limiting is discussed in the various flight control and flying qualities specifications and standards. However, the design and test techniques required to avoid major issues are not documented nearly as well as many of the other more fundamental aircraft characteristics such as short period frequency and damping, roll mode time constant, equivalent systems analysis techniques, etc. The specifications and standards acknowledge that excessive rate limiting can lead to major issues, but do not provide much specific methodology to help designers and testers determine the boundary between "excessive" and "adequate." Part of the problem has been that it is very hard to generalize the negative impacts of rate limiting in the context of specs and standards. Each project has had to come up with a design philosophy to avoid excessive rate limiting for that specific aircraft. Some projects developed and applied outstanding and unprecedented design philosophies in focused attempts to avoid rate limiting issues but still discovered major problems with rate limiting implications during flight tests. Despite the best efforts of design teams, unpredicted incidents with rate limiting as a contributing factor are likely to continue occurring on future projects. Flight test and design teams need to be better prepared for that eventuality as described in the Appropriate Action section below.

One of the reasons this topic has been a tricky issue for many decades is the uncertain boundary between: rate limiting that is considered "excessive" versus "adequate." One philosophy that has been promoted is to ensure that the rate limits are high enough and that the command gains and feedback gains are low enough that the control surfaces will never actually reach the rate limit value. That philosophy is probably impractical due to the fundamental tradeoffs between cost and intended system performance.

On relatively recent projects (within the last 2 to 3 decades), the more common types of rate limited oscillations that have slipped past the design process and have first been discovered in flight test have typically been associated with unusual flight conditions, unexpected "trigger" events, or other non-linear phenomenon. Those oscillation types tend to be harder to predict during the normal design process because models of aircraft components do not always reflect actual in-flight characteristics. See Lesson Learned "CA9: Update System Models with Flight Test Data."

The first "line-of-defense" against the occurrence of rate limited oscillations is an aggressive design philosophy to identify and eliminate the worst cases long before flight testing begins. When the design philosophy first line-of-defense is insufficient and these rate limited oscillations are still discovered in flight test, a well prepared test team is the second line-of-defense and needs to be focused on minimizing the potentially negative impacts. Those aspects are the main focus of this Lesson Learned as described in the Appropriate Action and Supplemental Discussion sections.

APPROPRIATE ACTION: Most airframe contractors do an excellent job designing the fundamental flight control system to meet requirements while also avoiding typical problem areas that could lead to major mishaps. Most designers are very familiar with the basic techniques involved for linear, low order systems and know how to adapt when the system requires higher order mechanizations since all of those techniques are well documented and an inherent part of any designer's experience. However, designer philosophies regarding the implications of control surface rate limiting can be much less consistent. Those philosophies have been observed to run the full gamut from "the design will be modified until any tendency towards rate limiting has been minimized" to "rate limiting is a normal part of controlled flight and people get too spun-up about the whole topic." That spectrum of design philosophies can exist even within the design community of a single airframe contractor organization. Part of the explanation for this trend is that the existing specifications and standards are much less explicit about the implications of rate limiting compared to many of the other aspects of flight control system design which are very well documented. Individual designers who are more complacent probably haven't yet experienced major mishaps with rate limiting implications.

Future flight testers on any given project need to understand designer philosophies regarding control surface rate limiting. If the designers demonstrate a healthy concern and implement a robust analysis, simulation, and flight test approach, special attempts to influence the process may not be required. If the designers exhibit excess complacency, more explicit discussions about rate limiting should be advocated within the relevant Integrated Product Team (IPT). If the IPT does not take appropriate action it may be necessary to elevate the topic through USAF technical representatives at the Program Office or on Safety Review Boards, Executive Independent Review Teams, and/or the Airworthiness authority.

Long before first flight of a new aircraft design, an important goal for flight testers should be to ensure that the analysis, simulation, and flight test approach includes explicit cases aimed at identifying occurrences of excessive rate limiting for further resolution. Frequency response analyses to identify the phase and gain changes due to sustained rate limiting should be part of the plan. Explicit simulation cases need to include dedicated setups that contain large amplitude control input reversals at flight conditions with high command and feedback gains, all in combination with external disturbances such as severe turbulence or jet wake encounters. An active literature search should be accomplished to identify the design and test techniques that have been applied on the most recent projects and also relevant historical projects.

Technical "home office" organizations within the Airworthiness community and at the Lead Developmental Test Organization should be able to assist with the literature search to identify the most relevant documents and also suggest focused analysis, design, and test techniques.

Once the design process has been assessed and/or influenced to provide the most robust process practical, the flight test team will still need to be prepared to assess in-flight occurrences of rate limiting. The best preparation is to have specific time history predictions for each planned flight test point. These predictions can be generated by batch simulations, piloted simulations, or a combination. Reviewing these predictions prior to each test mission is an invaluable step towards informing the specific test team members in the control room (both government and contractor) about what is considered "normal" regarding rate limiting for the aircraft/flight control system configurations, flight conditions, and maneuvers expected on that mission. See additional discussion about using predictions in the Supplemental Discussion section.

Besides being prepared with predictions, there are two main things that need to happen to help avoid potential mishaps. First, the discipline engineers in the control room need to be actively looking for sustained rate limiting during any pilot tasks and they need to communicate that fact to the pilot when it does occur. Since some amount of sustained rate limiting can be expected during almost any pilot task, a predefined, pre-coordinated criterion is needed to avoid a lot of "false alarms", or potentially worse… lack of an adequate warning from the control room. A technique that has been used very effectively to help the engineers in the control room understand if rate limiting is really occurring, is to have a dedicated plot showing surface rate. On a plot of surface rate, a case of rate limiting would show up as a fairly obvious square wave instead of the more ambiguous triangular-looking waveform on a time history plot of position. An alternative is to color code (or otherwise highlight) the portions of the surface position plot that are actually on the rate limit. Second, the pilot and test conductor need to be fully informed about the mishap potential associated with sustained rate limiting. It should not be assumed that experienced test pilots and multi-decade test conductors already understand the implications of rate limiting. The best way to ensure they are properly informed is to set up dedicated simulator demonstrations in which specific cases are selected with worst case examples of sustained rate limiting for that aircraft. Ideally the pilot and test conductor would be shown cases with and without rate limiting for similar tasks and flight conditions. If the baseline aircraft has been so well designed that it doesn't really have any worst case examples of rate limiting, it may be necessary to implement off-nominal aerodynamic models in the simulator to better demonstrate the insidious aspects of rate limiting.

The simulator cases with rate limiting should be configured with the goal of demonstrating how the aircraft responses don't track with the intended commands and how tightened pilot gains could exacerbate the situation. Once the pilot and test conductor have seen these potential mishap scenarios in the simulator, it will be easier to recognize and properly respond to similar scenarios during flight test. The simulator training should create an instinctive test pilot reaction to the undesirable aircraft response associated with a rate limited oscillation (i.e. a sense of added phase lag). The desired instinctive reaction would be to return to a previously cleared portion of the flight envelope. Normally, that same response works well for any oscillation. However, the test pilot should also be trained to avoid reversing large amplitude control inputs while attempting to return to the previously cleared envelope since those inputs could aggravate the rate limited situation. It would normally be better to choose a recovery direction and smoothly maintain inputs in that direction. Hence the need to show the pilot various worst-case scenarios in the simulator for the particular aircraft under test, even if those scenarios need to be somewhat contrived by implementing off-nominal aerodynamics.

The members of many test teams (both government and contractor) are notoriously reluctant to define a pre-coordinated continuation criterion because they "don't want their hands to be tied unnecessarily based on an arbitrary criterion". They are correct. The criterion should not be arbitrary… but it does need to exist. On numerous projects it has been amazing to watch how quickly those same reluctant team members can come up with a useful continuation criterion AFTER an in-flight incident has occurred. See Supplemental Discussion for examples.

The phrase "continuation criterion" has been intentionally used instead of "termination criterion." Both phrases amount to the same thing but have different psychological impacts. The phrase "continuation criterion" implies that if you meet this criterion you can continue. The phrase "termination criterion" sounds more negative because it implies that if you're just a bit outside the criterion you have to stop.

SUPPLEMENTAL DISCUSSION: Examples of rate limited oscillation incidents:

- *Flying Qualities Evaluation of the YF-16 Prototype Lightweight Fighter … a.k.a. "Flight Zero".* (AFFTC-TR-75-15, reference 10).
- *NASA F-8 Digital Fly-By-Wire Rate Limited Oscillation -Flight Testing at Edwards)*, Flight Test Engineers Stories, 1946-1975, reference 11).
- *YF-22A Engineering and Manufacturing Development Flight Test Summary* (AFFTC-TR-92-18, reference 12).
- *F-16C Block 40 Flying Qualities and Revised Pitch Integrator Rate Limits Flight Test Evaluation* (AFFTC-TR-96-24, reference 13).
- *F-22A Jet Wake Encounter Demonstration* (AFFTC-TR-06-16, reference 14).

Using Predictions to Better Prepare Test Team for Rate Limited Oscillations:

Predictions for many "open loop maneuver block" types of test points often exhibit very little, if any, tendency for rate limiting. Predictions for some test points (typically at lower airspeeds) may show a trend for a half cycle or possibly even a full cycle of sustained control surface rate limiting. If a certain amount of rate limiting was predicted for a given test point, there may be no reason to terminate testing if that same level is experienced in-flight. However, if the duration of sustained rate limiting experienced in flight is longer than predicted, the test team will need to have a pre-defined and pre-coordinated criterion in place to define "how much is too much."

Some projects have used "two full cycles of sustained rate limiting after pilot inputs have ceased" as a knock-it-off criterion for "open loop maneuver block" types of test points. This assumes that truly objectionable cases with two or more cycles were identified and "weeded out" during the design process and therefore not predicted. If predictions show two or more cycles of sustained rate limiting (after pilot inputs have ceased), it probably indicates a less than robust design process and those test points should be approached with extra caution.

The team members on some projects have decided that a more conservative rate limiting criterion was required and used "one and a half cycles" as the basis for a criterion. That more conservative criterion is especially appropriate when the test aircraft is in a configuration or at a flight condition where it has negative static margin (many modern aircraft fit in this category… and not just fighters). As the static margin becomes more negative, the criterion needs to be more conservative. A criterion like this has been used very effectively to guide real-time control room decisions about continuing envelope expansion or not. In most cases, post-flight analyses supported the decision to terminate envelope expansion for that flight and often resulted in a combined design and test team consensus to postpone envelope expansion until a control law revision was available. In rare cases, after the chosen criterion was exceeded and testing terminated, post-flight analyses were accomplished that supported a resumption of envelope expansion without a control law revision. However, in those cases test team awareness was heightened and continued envelope expansion was approached with a properly increased level of caution.

Predictions for many "closed loop piloted task" types of test points tend to be more problematic with respect to rate limiting. The problem is that many aircraft will exhibit sustained rate limiting if the pilot continues to apply large amplitude control reversals at an especially sensitive frequency. Large amplitude control reversals at very low frequencies will most likely exhibit rate limiting for a short duration after each reversal, but these cases typically do not exhibit sustained rate limiting after the pilot input has been relaxed. These cases also tend to be constrained by one or more aircraft structural limits or departure limits and therefore do not represent practical control inputs during piloted tasks. Large amplitude control reversals at very high frequencies are often "filtered out" by the control laws and do not exhibit sustained rate limiting. However, when the "sweet spot" of amplitude reversals and frequency occurs, it may be possible for the pilot to maintain the control surfaces on the rate limits for many cycles.

If the pilot applies control reversals in an "open loop" manner (i.e., without trying to accomplish any particular task) the primary result of the sustained rate limiting may be for the aircraft to wander off of the intended flight condition and eventually reach an aircraft limit. This might be the case during a pilot commanded frequency sweep intended to obtain in-flight data for frequency response analyses. Normally the control input amplitudes are kept relatively small during frequency sweeps in order to avoid non-linear effects on the data. However it may still be possible to see some sustained rate limiting during frequency sweeps even at those smaller amplitudes. That type of open loop sustained rate limiting wouldn't typically lead to a mishap situation. However, if the pilot suddenly switched to a closed loop task such as aggressively attempting to keep the wings level, the influence of the sustained rate limiting could cause the situation to quickly transform into a PIO as described in the following paragraphs.

Recurring Scenarios with Excessive Rate Limited Oscillations:

Without a doubt, the most insidious occurrences of sustained rate limiting have occurred during closed loop piloted tasks. These cases have exhibited very large amplitude PIOs that have exceeded aircraft limits and resulted in numerous Class A mishaps (see abbreviated list of examples at beginning of Supplemental Discussion section).

One of the reasons these cases tend to be so insidious is that the pilot typically doesn't know that rate limiting was occurring. From the pilot's perspective the aircraft responses simply weren't tracking with the commands. In these situations the pilot often feels like the intended task could be accomplished just by being more focused and trying a bit harder. However, the main closed-loop impacts when rate limiting is occurring are a significant increase in phase lag combined with a loss of overall command gain. The loss of command gain tends to make the pilot want to try harder while the increase in phase lag tends to make it much harder to accomplish a closed loop task without resulting in a large amplitude PIO. That combination contributes to the pilot using "bang-bang" control inputs (meaning repeatedly applying full stick in one direction followed by full stick in the opposite direction) in an attempt to control the oscillations. If the engineers in the control room do not inform the pilot that rate limiting was occurring the pilot will probably want to repeat the test point. If the point is repeated while the pilot is "trying a bit harder" that can also be interpreted as "higher pilot gain" which can easily exacerbate the problem and lead to higher amplitude rate limited oscillations that rapidly degrade into a potential mishap.

Another reason these cases tend to be so insidious is that when the control surfaces in a given axis are "stuck" on the rate limit due to pilot inputs, the feedback paths in that axis are effectively disabled for the duration. Therefore any stabilizing influence of those feedback paths no longer exists. Instead of flying a well-behaved augmented airplane, the pilot has to deal with the much poorer handling qualities of the unaugmented aircraft. Similarly, if the control surfaces are "stuck" on the rate limit because of elevated gains in the feedback paths, the pilot inputs will no longer be as effective as expected since the surfaces may or may not respond in the correct direction.

In a CAS-type design, the command gains may have been raised to high levels in an attempt to achieve increased agility not possible with an unaugmented flight control system. In these situations, relatively small pilot inputs could drive the actuators onto rate limits. Typically this would not be perceived as a problem if the pilot inputs were applied in only one direction such as a "stick snatch." In those situations the control surfaces would only be on the rate limit for a very short time. However, if the pilot were to reverse control inputs for whatever reason the control surfaces could be on the rate limit in the original direction, and then go onto the rate limit in the opposite direction. There can be a variety of "triggers" to induce the pilot to apply large inputs in one direction and then reverse those inputs. In this example, the only "feedback loop" may be through the pilot (i.e., no other sensors feeding back through the flight control system). The situation can be greatly exacerbated if the pilot senses the aircraft is not properly responding to control inputs. That overall scenario has been one of the classic setups for a large amplitude PIO.

The increased likelihood of PIOs when control surfaces were sustained on rate limits has been attributed to increased phase lag compared to the baseline system (without sustained rate limiting). Recent large scale EMD acquisition projects have applied a great deal of design and analysis effort to forecast and minimize the potential for rate limited PIOs. Nevertheless, a number of cases have "slipped through the cracks" and have resulted in Class A and Class B mishaps.

In a SAS-type design, the feedback gains may have been raised to high levels to improve the pilot perception of closed-loop stability and overall handling qualities. When this method is applied to aircraft that are statically or dynamically unstable the results can provide greatly improved handling compared to the unaugmented airframe. However, the more inherently unstable the aircraft, the harder the flight control system has to work to achieve the intended result. When the feedback gains become high enough to easily force the surfaces onto sustained rate limiting, unexpected results can happen suddenly and violently. If the pilot attempts to control the resulting large amplitude aircraft motions, the situation can be aggravated if the pilot inputs are in a direction to maintain the surfaces on the rate limits even longer than would have occurred without pilot inputs.

Rate limited oscillations have also occurred when the elevated feedback gains only apply in certain situations such as when operating near a g-limiter or angle-of-attack limiter. Rigorous and methodical simulations can weed out the most extreme cases, but the flight test team must remain vigilant for any unfiltered cases.

Flight test teams must also be very conscious of where the aircraft is in the flight envelope when rate limited oscillations occur. Sustained rate limited oscillations tend to be more common at lower airspeeds. These oscillations tend to be sustained at larger amplitudes and at lower frequencies. However, the aircraft responses also tend to be slower so it may take several cycles before the situation degrades to the point where aircraft limits are exceeded. That time gives the pilot a chance to take corrective action. In some cases, large amplitude rate limited oscillations can be "self-correcting" because the large amplitude motions change the flight conditions and therefore change the command and feedback gains in a direction that may help avoid the rate limits. However, no test team should count on that extremely fortunate situation.

It can be much more serious if rate limited oscillations occur in the high speed portion of an aircraft flight envelope. These oscillations tend to be sustained at lower amplitudes but at much higher frequencies. These situations are more serious because they can degrade much more rapidly. At these higher airspeeds it may not take much "unaugmented" control surface motion to drive the aircraft past structural or departure limits. The pilot may have insufficient time to apply the correct recovery controls, and those controls may not be as effective as expected because of the sustained rate limits.

Rate Limited Oscillations not Limited to Fighter Aircraft:

There is a common perception that fighter aircraft are more susceptible to rate limited oscillations than larger aircraft such as transports or bombers. Some of the more “infamous” cases may have indeed been on fighter aircraft, but there have also been a surprising number of rate limited oscillation events on the larger aircraft. Part of the reason for this is that larger aircraft tend to have lower control surface rate limits. For example, a typical rate limit for a fighter aircraft might be about 50 to 60 degrees per second whereas a bomber/transport may have a rate limit as low as about 15 degrees per second. When those low rate limits are combined with a high gain digital flight control system, the potential for rate limited oscillations can occur at surprising flight conditions. For example, one large aircraft exhibited intermittent but recurring rate limiting while in contact with the aerial refueling tanker. This occurred at a routine aerial refueling airspeed for that aircraft and resulted in a Class A mishap with significant damage to the refueling boom on the tanker. Large aircraft may have scheduled rate limits that reduce the values at low airspeeds in order to minimize loads on those control surfaces and increase fatigue life.

FF10: PLAN FOR INDEPENDENT THIRD GENERATION DATA ANALYSIS TOOLS

LESSON LEARNED: Using data analysis tools developed by others outside the 412 TW undermines the veracity of the "independent" analysis and resulting conclusions and recommendations put forward by the 412 TW. It also could undermine the possibility of catching issues and problems that may be uncovered by analysis tools developed independently by experienced testers.

PROBLEM: Some test programs have attempted to cut costs by having one set of analysis tools, usually created by the developing contractor. The flawed thinking is that all data analysis is the same, so there is no point in duplicating efforts. This approach misses the point that both the developing contractor and the USAF testers bring very different but very valuable focuses in the types and thrusts of their respective analyses. It also creates a difficult position for the 412 TW to render "independent" conclusions and recommendations from analysis tools that the 412 TW can't corroborate.

DISCUSSION: The credibility of the 412 TW is only as good as the credibility of its results that are backed by sound substantiation. Many times the TW results are "bad news" to the program. When this occurs, the analyses and results are put under intense scrutiny. The veracity of the substantiating data and analyses is the key to withstanding such scrutiny and maintaining a reputation of credibility. On the flip side, it would be costly and possibly dangerous if the TW analyses and evaluation missed a problem that went to the field, which would also undermine the TW credibility. The primary means for the TW to uphold its credibility and have the necessary confidence to convey both the good news and bad news is for the TW to use its years of expertise in developing its own independent third generation analysis tools and do the analysis and evaluation that substantiates the results. Relying on analysis tools developed by others does not provide a basis for such confidence and credibility.

For the purpose of this Lesson Learned, some generic definitions for first, second and third generation data processing are described in the Supplemental Discussion section. Flight test programs have routinely shared the first and second generation data processing tools without much issue. The focus of these processing and analysis tools are the same for both the contractor and the USAF testers. There is no reason to believe this practice shouldn't continue. When it comes to third generation data analysis tools, most programs have had separate tools developed and used by the developing contractor, and by the USAF DT testers. Each brings a very valuable and very different perspective to their analysis tools. This approach has provided a more thorough assessment of the weapon system. The developing contractor focuses on their contractual obligations, such as spec compliance, model verification, and confirming their design considerations. The USAF DT testers also consider such things as spec compliance, but their analysis focus also employs insights and lessons learned from numerous other similar weapon system flight test programs. The insights and lessons learned imbued in the independent analysis tools developed by the USAF DT testers have frequently provided detection of issues not uncovered by the contractor's analysis, and/or provided details on problems that enabled fixes when the contractor's analysis was lacking such detail. See examples in Supplemental Discussion section. In addition, USAF DT testers have an obligation to evaluate overall Military Utility which is an aspect often lacking in contractor reports that tend to focus on individual subsystems and very detailed deliverables as specified in the contract.

There have been some test programs in which the Program Office and/or developing contractor have insisted on the team using one set of analysis tools, typically developed by the contractor. It was often done as a cost cutting measure, because independent tools were considered duplicative, or because they just wanted to try and avoid getting conflicting answers. This point of view was typically driven by managers with little to no first-hand experience with third generation data analyses from flight test data. The perception was that all data analysis should be the same regardless of the role of the people doing the analysis. This perception didn't recognize the very different focus of the developing contractor and the USAF DT testers. Both are necessary and valid, but do present substantially different considerations in the analyses and evaluations. The differences in motivations for analysis tools developed by contractor versus government developmental testers is described below in the Supplemental Discussion section. Hopefully an understanding of those differences in motivation will help future engineers counteract the typical concerns that have led to inappropriate restrictions on their ability to accomplish their job.

In addition, any cost savings accrued by precluding development of independent analysis tools was almost immeasurable compared to the daily costs of the flight testing that collected the data. Therefore, avoiding independent analysis tools and independent analyses did not maximize the return on the investment in collecting the data. Another problematic aspect of this approach was that the USAF testers were not allowed to review the code in the contractor developed analysis tools to determine the veracity of what was being done to the data, based on claims of proprietary intellectual capital. There have even been extreme instances of the contractor manipulating the analysis tools so the results mimic preconceived notions (see example 1 in Supplemental Discussion section).

Some programs that pushed for one set of analysis tools attempted to placate the USAF DT tester's concerns by claiming the tools would include analysis elements requested by the USAF testers. Experience has shown this doesn't work well. The contractor was responsible for implementing the USAF requirements, but when the inevitable schedule issues arose, the USAF requirements weren't met, and the contractor claimed large cost overruns if forced to meet the USAF testers needs. In addition, the USAF testers weren't allowed to review the code to determine its veracity, based on the proprietary intellectual capital issue described above.

APPROPRIATE ACTION: In the early stages of a program, the USAF DT testers should lay out a plan to have independent third generation data analysis tools. Consideration should be given to existing proven tools from within the 412 TW and any new tools needing to be developed within the TW. This should be included as an integral part of the engineering man-hours estimate provided in the Statement of Capability. Part of those estimates can and should include a tradeoff showing that it will be cheaper to use existing 412 TW tools than it will be to spend a lot of combined government and contractor time to adapt to alternative methods with questionable benefits. If there is direction from the Program Office to have one set of analysis tools developed by the contractor for use by all on the team, work with them to develop an understanding of the benefits of the independent analysis tools, and the relative return on the investment, pointing out that the cost is minuscule compared to the cost of collecting the data. If the pressure continues, elevate the issue within the TW, to highlight the fact that the TW is being put into a potentially difficult position to render credible and complete results, conclusions and recommendations that the TW can substantiate.

It may help alleviate concerns at the Program Office regarding duplicate analyses and potentially conflicting results by agreeing to common analysis tools for first and second generation processing (as discussed in the Discussion section above and described below in the Supplemental Discussion section). Those are the areas of data processing most likely to lead to conflicting results due to minor differences in unit conversions or the choice of a specific calculation method. It is appropriate to attempt to find common ground for those conversions and calculations since they just tend to reflect the most basic and necessary standard equations for all users. On the other hand, the government analysis team NEEDS to understand the specific choices made for third generation analyses. That has often not been allowed when the specifics of those equations and tools were considered proprietary. The overall best balance between common analysis tools and independent roles may be achieved by "drawing the line" at third generation tools as discussed below.

SUPPLEMENTAL DISCUSSION: Each project will probably come up with its own definitions for the boundaries between first, second, and third generation data processing. To help explain the concepts outlined in this Lesson Learned, the following generic definitions are used:

First Generation Data Processing: This generation is focused on obtaining basic engineering unit data from the raw measurements. Those raw measurements have typically been obtained as a voltage from analog instrumentation sensors or as digital 1553 MUX data from either production data sources or for flight test specific “data pump” variables. More recent projects have needed to obtain test data from multiple types of MUX Bus sources (not just 1553), and also from Ethernet-based sources. The various types of data sources will undoubtedly continue to grow, but the process remains fairly straightforward and is well suited to sharing data processing by both the contractor and government stakeholders. The contractor is typically expert at knowing the correct calibrations for all types of sources (although a little cross-checking from government flight testers has proven to be advantageous on multiple projects).

Second Generation Data Processing: This generation starts with basic engineering unit data and applies standard, fundamental calculations to provide more useful data as inputs to more complex data processing. Examples include: correcting data from the sensor location to the center of gravity or pilot station, adjusting for test day temperatures that did not match a standard-day atmospheric model, time aligning data from multiple sources sampled at different times and data rates, smoothing data to reduce the impact of sensor noise or atmospheric turbulence, etc. This type of data processing tends to be relatively straightforward and is also well suited to sharing data processing by both the contractor and government stakeholders. In order to achieve peaceful coexistence, each organization may have to compromise on its favorite smoothing algorithm or time alignment technique in order to obtain a result that is adequate for all users.

Third Generation Data Processing: This generation starts with the basic data that has been modified by the fundamental and necessary corrections. At that point, the ensuing data analyses can go in many different directions based on the motivations and experience level of the analysis teams. The contractor is highly motivated to show that the product meets contractual requirements. The government DT testers are motivated to get deficiencies addressed and to ensure that the final product will meet the needs of the operational users. Those diverging motivations will influence the type of data processing selected.

Examples of typical contractor-oriented analysis tools include:

- Showing compliance with very explicit specification or Airworthiness requirements,
- Demonstrating compliance with pre-defined incentive or award fee criteria,
- Verifying that existing system models adequately reflect flight test results or updating those models to be “close enough” to flight test results.

Examples of typical government DT-oriented analysis tools include:

- Summarizing the results from dozens or hundreds of flight test missions to “paint an overall picture” of the actual capabilities of the final product,
- Capture the qualitative opinions from multiple test pilots to reflect an overall assessment of aspects that are ready to proceed to Operational Testing or that may require further improvement (while also attempting to backup those qualitative opinions with quantitative data),
- Providing an assessment of Military Utility aspects of importance to the Operational User
- Highlighting aspects of system models that do not adequately reflect flight test data

Both government developmental testers and on-site contractor engineers share a strong focus on flight test-oriented data analysis aspects such as pre-flight predictions, flight-to-flight clearances, and resolution of flight test issues. However, on-site contractor engineers tend to operate more as “conduits” to get data back to home office decision makers, and those decision makers have often been much more focused on priorities other than flight test. Government developmental testers tend to have a much broader data analysis scope than their on-site contractor counterparts.

The diverging motivations that inspire differences in third generation analysis tools and products make it highly unlikely that some combination of methodology will satisfy the needs of both contractor and government DT testers. Therefore management should not force all users to utilize the exact same process. If common ground can be found for some specific third generation analysis tools (such as aerodynamic Parameter Identification as described in example 3 below) some programmatic efficiencies may be obtained, but it is more likely that attempting to achieve a “one-size-fits-all” third generation analysis process will be a detriment to overall efficiency for all stakeholders.

Here are some real examples of issues resulting from having to use analysis tools developed by the contractor, and tools the contractor failed to provide.

Example 1: On a major aircraft development program, the System Program Office insisted that the contractor’s data analysis tools would suffice, and therefore the entire DT test team was directed to use them. The CTF capitulated to avoid conflict with their customer. Late into the performance testing phase, the USAF DT test team received analyses from the contractor addressing the range and fuel flow characteristics. The data reflected nearly perfect matches with the predicted data, which was quite unusual in all areas of the envelope. The USAF testers generated their own analyses from the first and second generation data in segments of the envelope, which showed significant differences from the predicted performance, and contractor’s results. After much contentious discussion, it was discovered that the contractor could not believe their drag predictions were off, so they manipulated the position error and airspeed corrections to get the predicted performance. All remaining contractor produced analyses became suspect, so the USAF DT testers developed independent analysis tools in the Flight Dynamics disciplines. Additional issues were found in structures and flight controls which were not uncovered through the contractor’s analysis tools. One such issue was a severe flight control instability that would have been catastrophic if that part of the system was activated in flight. The contractor’s analysis tools did not identify the looming instability.

Example 2: On the same program discussed in example 1, the aircraft frequently experienced nose gear chatter when landing/braking and on some occasions when taxiing and braking. At times, the chatter became so intense, the pilots had difficulty keeping their feet on the rudder/brake pedals. The contractor did not have the data analysis tools that could isolate the cause and effect to design a fix. The USAF DT testers stepped in and performed a frequency response analysis on the data using data analysis tools developed within the 412 TW based on years of experience. The cause and effect were isolated, enabling a fix that eliminated the problem.

Example 3: On another major aircraft development program, the contractor design and test teams did not have much experience with flight test Parameter Identification (identifying specific elements for aerodynamic models based on flight test data… as opposed to relying exclusively on data from wind tunnel testing). On the other hand, some of the government DT testers were extremely experienced with flight test Parameter Identification. Those testers were able to isolate the specific unpredicted root cause for a very significant flight test anomaly and provide corrected aerodynamic models that were eventually used by the contractor to implement design fixes into the flight control system. Although the example described here was for a particular development project, the same words could be used to describe a dozen or more development projects with similar situations. The Parameter Identification experience levels seem to ebb and flow constantly at any given contractor. Some contractors have had industry recognized Parameter Identification experts whereas others have had no relevant background whatsoever.

Example 4: Some of the best projects have had a very limited number of true Parameter Identification experts on both the contractor and government test and analysis teams, but were able to share the workload by modeling different portions of the flight envelope. In that case the contractor engineers analyzed the majority of the flight envelope at low angles of attack, whereas the government engineers were able to accurately update the aerodynamic models for the tricky high angle of attack region. That particular situation did not reflect the original plan but evolved out of necessity as skilled personnel were difficult to find for either organization.

FF11: CROSSWIND TAKEOFF AND LANDING ENVELOPE CLEARANCE

LESSON LEARNED: On many large scale EMD-type development projects, flight testing to clear crosswind takeoffs and landings to the maximum required limits has tended to be delayed until the end of major development (such as EMD) or sometimes postponed into a follow-on test effort. Many factors have contributed to this trend but the end result has been that operational users have been required to get by within partial crosswind limits for a long time. This lesson suggests several methods to minimize the negative impacts and hopefully result in more timely clearance to the full crosswind limits.

PROBLEM: Fully clearing the crosswind takeoff and landing envelope has been a recurring challenge on almost every major EMD-type development project. It has been fairly common for the crosswind envelope to be cleared reasonably quickly out to an interim value such as 15 or 20 kts of direct crosswind. Those interim limits are usually adequate to allow ongoing flight testing without excessive impacts. However, testing out to the maximum limits (often around 30 kts) has typically been constrained by a host of natural and man-made phenomenon that resulted in long delays for full envelope clearance. Those delays have sometimes stretched out long after Initial Operational Capability and impacted the ability of operational squadrons to generate sorties.

DISCUSSION: There are numerous examples of situations on past projects which contributed to lengthy delays prior to full clearance of crosswind takeoffs and landings. Some of those situations are considered “normal” for crosswind testing, but the negative impacts could have been avoided or improved by better understanding during preparation. Many of the situations encountered were somewhat “artificial” and could have been improved by avoiding unrealistic assumptions. Sometimes these artificial impediments were not instigated by the local test community but were the result of decisions and test philosophy established by external people and organizations. The Supplemental Discussion section provides more background information for both the “normal” and “artificial” situations. The proposed solutions for these situations may seem obvious, but the negative impacts of these situations have been consistently underestimated on multiple development projects.

In addition to the various normal and artificial situations described in the Supplemental Discussion section, “Loss of Test Expertise” is another aspect that will undoubtedly impact future test projects. There is no better preparation for crosswind testing on a new aircraft than having recently completed a similar project on another aircraft. However, new aircraft designs will show up at the test community less frequently, and crosswind envelope expansion will only encompass a small portion of each overall project. Each project will need to take stock of the specific experience of project personnel to evaluate readiness, and hopefully recognize when additional help is necessary.

APPROPRIATE ACTION: A very short summary to capture the “essence” of the Supplemental Discussion is to:

Implement Realistic Plans for “Normal” Crosswind Testing Situations:

- Do not commit to using the Edwards AFB (EAFB) main runway as the only practical option
- Do not depend on using the EAFB lakebed runways (especially if the test aircraft has special coatings on the exterior surface or engine inlet locations that increase the potential for foreign object damage)
- Test plans should recognize the practical aspects of testing at or near the aircraft design limits
- When it is windy, the entire team must be ready on short notice to go execute crosswind testing at the appropriate location

Avoid Unrealistic Assumptions that lead to “Artificial” Impacts on Crosswind Testing:

- Aggressively advocate a higher priority for crosswind testing when there is still time left before the end of the overall test program
- Take advantage of early tester involvement to ensure that the instrumentation is designed to be able to support crosswind testing without significant reconfiguration and/or downtime.
- After recognizing EAFB runways will probably not support all required crosswind testing, ensure adequate coordination accomplished for the alternative test sites
- Take full advantage of on-board GPS/INS capability for real time wind calculations to reduce dependence on portable weather stations next to the intended runway
- When accomplishing crosswind testing at remote locations, be prepared for loss of telemetry at EAFB (when the test aircraft gets down to low altitudes at the alternative location) by having the appropriate “re-radiation” capability or portable mission control room already in place
- Avoid simultaneous tight tolerances in multiple dimensions

SUPPLEMENTAL DISCUSSION:

“Normal” Factors that Influence Crosswind Takeoff and Landing Testing.

EAFB Main Runway Alignment: Edwards AFB has a very long and wide main runway (roughly 15,000 feet long and 300 feet wide). That concrete runway is very suitable for crosswind landing testing but it was designed to align with the prevailing wind direction (roughly from the southwest). Direct crosswind components of 15 to 20 knots do occur, but tend to be infrequent. Direct crosswind components of 30 knots are unusual. Although total winds of 30 knots or higher are common, those high winds tend to blow almost directly down the main runway (as intended). On those rare occasions when the direct crosswind components are high, the variation in gust levels also tends to be high, which can lead to exceedance of landing gear limits on individual test runs. Edwards AFB now has an alternate runway, but it is aligned in the same direction as the main runway. Therefore another runway option (other than the EAFB main runway) is typically needed in order to complete crosswind takeoff and landing testing out to the maximum limits for a given airframe. The runways at Mojave airport, Mojave, California, have recently been used successfully. The runways at Victorville and Palmdale, California, may also be worth consideration if they meet the needs of the intended crosswind testing.

EAFB Lakebed Runways: Historically, the EAFB dry lakebed runways have provided plenty of options for accomplishing crosswind landing testing. There are multiple runways of various lengths, widths, and alignments relative to the prevailing wind. A Compass Rose on the lakebed has also been used for low approaches at any desired angle relative to the local wind, but that feature may not be maintained without proper pre-coordination with Airfield Management. However, actual landings or touch-and-go patterns on the lakebed are not compatible with modern low observable aircraft. The many small particles and larger rocks kicked up by either the main or nose gear can cause considerable damage to the underside of the aircraft external surfaces and can lead to lengthy and expensive repairs. Therefore the lakebed runways are not suitable for crosswind testing on those types of aircraft. Lakebed testing may also be inappropriate for test aircraft with low engine inlets due to the increased potential for foreign object damage.

Greater Difficulty near Crosswind Limits: It should not be a surprise that testing to the crosswind design limits of a given aircraft can be extremely challenging. However, many projects have still been caught off guard by that level of difficulty and been inadequately prepared. Testing up to 15 or 20 knots of direct crosswind can be tricky, but testing at crosswind design limits such as 30 knots brings a number of additional critical issues into play. Chief among those issues is the fact that it becomes more critical to precisely control test conditions (to avoid exceeding gear load limits) at the same time that it becomes impractical to do so (because of gusty wind conditions). In order to avoid exceeding landing gear loads, a variety of variables must be carefully controlled. Some of those variables include: sink rate at touchdown, side load on the landing gear, skidding effects on the tires, and the implications of handling qualities and engine responsiveness on the ability of the pilot to control touchdown conditions. Unfortunately, it is very unusual to get nice steady direct crosswind components at exactly the target value. It is much more common to have very large gusts superimposed on the steady portion of crosswinds, with the overshoots increasing the potential to greatly exceed the design limits. The most successful projects have recognized the higher level of difficulty and had plans in place to account for those issues. An important example is to recognize and account for the ability of the landing gear to accept overshoots in side force at touchdown. If the test aircraft landing gear is likely to be considered overloaded if touchdown occurs at 32 knots of direct crosswind, it is unreasonable to expect a test team to control conditions well enough to accomplish a test point that requires touchdown in a very tight band such as 25 to 30 knots direct crosswind. It has been surprisingly common for crosswind landing test points to be defined such that the actual test conditions must be accomplished within a combination of extremely tight tolerances (i.e., the test point will not be considered complete until the test results are within 25 to 30 knots direct crosswind, while also keeping sink rate within 1 to 2 feet/sec, while also within 1 or 2 knots of the target airspeed, etc.). At the design crosswind limits it may be possible to meet all of those tight test tolerances at the same time, but the end result could depend more on blind luck than proper test planning. It is much more likely that the combination of high programmatic pressures, tight tolerances, and low overshoot constraints will set the test team up for an event which causes structural damage and is considered a mishap. It may require heroic efforts to avoid that type of predictable mishap. Some landing gear designs may be considered especially robust with lots of overshoot capability. Some landing gear implementations have been "off-the-shelf" designs that had been used successfully on another aircraft, but may not provide much margin for the new aircraft implementation. It is common for the landing gear to be designed (or selected) with just enough margin to barely meet design requirements. In that case the test team should recognize the importance of that constraint and insist on defining realistically broad test tolerances. That can be extremely difficult when the fundamental test requirements are being defined by multiple off-site organizations, but may be necessary for test safety. The same concepts described above may also be necessary and effective to address overlap with the artificial situations listed below.

"Artificial" Impediments that Can Impact Crosswind Takeoff and Landing Testing.

Perceived Test Priority: Crosswind takeoff and landing testing tends to have lower priority than many other types of testing that have more immediate operational or programmatic implications. However, when crosswind testing has been consistently postponed in favor of higher priorities, the end of the development program has become imminent but without a fully cleared crosswind envelope. In some cases crosswind testing has been one of the few remaining flight test options but in seasons when high crosswinds were especially rare. In those situations program management saw no option but to declare the basic development program complete, and defer full crosswind clearance to a follow-on effort. The artificial impact of previous decisions regarding test priority delayed the option for conducting crosswind testing until it became impractical.

There have been other impacts of delaying the lower priority crosswind testing until late in the development project. When the remaining time frame is short, the test team tends to get very focused on accomplishing the final test points (normally considered an admirable trait), but the team may not adequately consider the practical aspects of those very tricky end points. For example, during crosswind landing testing approaching the final limits near the end of development, at least one project has had a main gear tire blowout due to MUCH higher tire wear than anticipated. The tread on the tire eroded much more quickly at the higher crosswinds than prior experience (at roughly one third the normal number of landings before tire replacement). Future test teams need to be very conscious of this trend and build test plans to include the appropriate inspections at much higher frequencies than for less challenging crosswinds. Those inspections may need to occur after only a handful of crosswind landings because of the scrubbing that occurs when the crab angle is not fully removed prior to touchdown. This may limit a single flight to just a few touch and go landings at high crosswind conditions. A constraint like that may be necessary, but can be very frustrating when the test team had waited months for the right combination of factors to enable those test points to be accomplished.

Instrumentation Configuration: It is quite common for crosswind takeoff and landing testing to require special instrumentation, typically for landing gear loads. Some projects have chosen to implement that instrumentation configuration as a "package" that must be "swapped" between the configuration that supports crosswind testing and a separate configuration that supports alternative types of testing. On those projects it could take two weeks or longer to accomplish the instrumentation swap. The unfortunate (and foreseeable) result has occurred when the test aircraft had spent two weeks of downtime to install the crosswind testing instrumentation package but the winds required to support crosswind testing did not occur. In those situations, the test aircraft could not support the alternative testing because the instrumentation was not configured as needed. Therefore the test aircraft stayed on the ground, unable to accomplish any productive flight testing. After another two week downtime to swap back to the alternative configuration, the crosswinds would inevitably increase to testable levels, but with the wrong instrumentation package installed. The artificial constraint imposed by the cumbersome instrumentation design was a severe impediment to successful crosswind testing.

Test Preparation Constraints: Some projects have belatedly recognized that crosswind testing would not be practical using either the EAFB main runway or lakebed runways. The logical decision was eventually made to use a runway with better alignment relative to prevailing winds at another location. Several choices are available within range of EAFB without the need for inflight refueling (such as Mojave, Victorville, and Palmdale). However, testing at those other locations induced additional coordination requirements that further delayed testing. Those additional requirements have included things like: ground access near the runway (for a portable wind station, portable theodolite, or similar device), arranging for local firefighting support (in the event of hot brakes or other mishaps), or security support and ground access for maintenance personnel (in the event the test aircraft needed to land at the remote site). Artificial constraints were inflicted by not planning for those foreseeable test requirements, leading to the additional delays.

Crosswind Measurement Constraints: Some projects have had to deal with unnecessarily onerous requirements for measuring crosswinds which have impacted the ability to execute crosswind testing. Sometimes these onerous test requirements have been self-imposed from within the project. Sometimes they have been imposed by external entities during the review process. One recurring example has been the requirement for a portable weather station onsite at the test location to measure local winds. While a portable weather station for wind measurement may seem like "a good idea", that does not mean it should be legislated as a hard requirement for either safety calls or technical data. Even though some remote locations may be within 20 minutes by air, it can easily take two hours or more to reach via surface roads. On the day of test, there have been numerous reasons that have kept the portable wind station from being onsite at the remote location and capable of supporting testing. As one example, the crew operating the portable wind station has been deployed to one remote location based on a wind prediction, only to find out later that the actual winds were much more compatible with test requirements at a different location. Fully capable test aircraft have been forced to return to base simply because the portable weather station was unavailable for "knock-it-off" calls as legislated in the safety package requirements. Although a portable wind measurement may be somewhat helpful to provide improved post-flight data, it is highly questionable as a source for real-time safety of test decisions. A portable wind station only measures wind at the specific spot it is located, which may be thousands of feet from where the test aircraft touches down or lifts off. One practical alternative has been to simply use tower winds for safety of test decisions (as would be done during normal operations). A better alternative is to use real-time, on-board winds from the test aircraft. Modern test aircraft tend to have very capable blended GPS/inertial navigation systems that can very accurately provide wind values at the test aircraft location throughout most of a takeoff, touch and go, or approach and landing. The accuracy of those onboard wind estimates can degrade during dynamic maneuvers or when sideslip values become large. However, it has become a straightforward process to discount inflight calculated wind estimates during those conditions and rely on the values calculated during the more stabilized portions of a test sequence. The primary obstacle to relying on the onboard GPS/INS wind estimates has been the perception that the process is "too complicated" to be trustworthy. To be effective, the onboard wind estimates rely on proper understanding of the test aircraft air data system during runway operations, especially when in ground effect. That understanding should already be part of a well-planned test project. In some cases the onboard wind estimates may need to be supplemented by additional corrections in the real-time displays in the mission control room. Once properly established as part of the real time data flow, blended GPS/INS data combined with confident air data calibrations should provide more than enough accuracy to support safety of test decisions and have been considered much more useful than a ground-based wind station. Once a test team has sufficient confidence in utilizing an onboard GPS/INS as the source for real time wind data, a portable wind station requirement can be eliminated or relegated to an "As Available" basis, not legislated as a hard requirement for safety or technical purposes.

Loss of Telemetry near Ground at Remote Locations: It is typical that engineers in the control room must be monitoring gear loads and flight control system reactions during crosswind testing. However, it has been surprising that some projects have been caught off-guard when real time telemetry from the test aircraft becomes very intermittent or is lost completely in the mission control room at EAFB when the test aircraft gets close to the runway at the remote location. Since that loss of TM tends to occur within the last 200 ft AGL, it can be very disruptive to real time decision making. In one case, the loss of TM during a remote crosswind touchdown caused the control room team to miss an overload condition that was not discovered until post-flight data review. One common solution has been to establish a "re-radiation" capability to capture real time data from the test aircraft and retransmit that data to the mission control room. Those "re-radiation" systems tend to be located at an intermediate location (sometimes on hilltops) with direct line of sight to both the remote runway and to the EAFB antenna system that feeds the mission control rooms. Another potential solution that has been underutilized is to use a mobile control room in the vicinity of the remote location. At various points in time EAFB has had a functional mobile control room. Many contractors have similar capabilities. However, those facilities tend to be viewed as an unnecessary expense by the business-oriented portions of many test organizations. Unfortunately, the desire to avoid that relatively small perceived expense has led to much larger expenses related to inefficient crosswind testing.

Tight Tolerances in Multiple Dimensions: One of the recurring themes that has contributed to the difficulty of completing crosswind testing has been when very tight tolerances have been defined on multiple factors that must all be met simultaneously. Sometimes this has been necessary to ensure that adequate coverage has been achieved and to obtain the essential data needed to provide a confident clearance for the end user. However, that combination has also created situations where the intended test point is almost impossible to achieve in any practical way. Therefore flexibility needs to be emphasized when defining a crosswind test point matrix. Examples of that flexibility might be in terms of: weight, CG, airfields, loading configuration, wind into heavy versus light wing, etc. If tight tolerances truly are required in some dimensions, it might be prudent to relax tolerances in other dimensions. The end result should be a carefully thought out test approach with serious consideration of the tradeoffs. Unfortunately, many past projects have not figured that out until it was too late.

CONCLUDING THOUGHTS:

The authors do not claim that the methods described are the only way to go. Other methods may work as well or better. The overall goal was to encourage future flight testers to contemplate each topic in the context of a given project and develop creative solutions for that situation.

There was no intent to "bash" any particular organizations or individuals from past projects. There is no doubt that the decisions implemented on those previous projects were made with the best intentions, but possibly without a full understanding of the eventual consequences. It is hoped that the lessons captured in this handbook will provide some insights into those eventual consequences, and may help avoid similar situations in the future. By working to create a genuine team with "foresight and willingness to act", it may be possible to avoid the "endless repetition" of flight test history.

As discussed previously, this handbook is envisioned as a living compilation that should be expanded in future versions to include lessons that have not yet been captured, encompass other technical disciplines, and the viewpoints of other flight testers. Some suggestions are made below for possible standalone handbooks and individual topics for consideration.

Possible Future Handbooks:

Envelope Expansion Lesson Learned – Volume 2

Mission Systems Lessons Learned

- Possibly with several volumes for specific test types and classification levels

Flight Test Operations Lessons Learned

- Possibly including viewpoints of test pilots, test conductors, and test directors

Flight Test Logistics Lessons Learned

- Possibly including the entire spectrum of maintainer activities from initial coordination and planning, to the assembly line, to mission support, to Tech Order Verification and Validation, to Suitability Evaluations

Possible Future Topics:

- How Developmental Testers can best work with Operational Testers
    - Some projects have had outstanding interactions between DT and OT, other projects… not so much. The techniques that have worked best need to be documented for the benefit of future programs.
- Writing and Using Interim Aircraft Operating Limits
    - Although contractors tend to be the primary authors of documents establishing fundamental interim limitations (i.e., before achieving the final Flight Manual limits), flight testers are the primary users of those documents. This topic should not be viewed as repeating directives from various regulations and instructions, but should focus on the practical aspects that most impact daily life in a flight test organization.

- How to function best in a Mission Control Room
  - Existing Crew Resource Management courses encourage people to "speak up" in a mission control room, and each Combined Test Force establishes its own training process. However, there are undoubtedly lessons that can still be captured from many previous projects on what works well and what doesn't.
- Pre-First Flight Sensitivity Studies
  - Primarily for off-nominal aerodynamics and air data variations, but may also be applicable to other technical aspects. Intended to help assess the contractor's process, but also to suggest techniques to influence that process if necessary.
- Assessing the Impacts of Mispredicted Aerodynamics
  - What to do when flight testing doesn't reflect the predictions. Summarize flight test analysis techniques that have worked well on previous projects. Capture positive programmatic impacts from those projects.
- Pilot In-The-Loop Oscillations (PIOs)
  - Capture existing reports, handbooks, and design documentation in a "one-stop-shopping" location. Also… to capture techniques for recognizing incipient PIOs during test missions before those situations degrade into a mishap.
- Recognizing "Un-obtainium"
  - Multiple projects have had super-high priority test points that were absolutely essential to meet award fee criteria, congressional milestones, or other critical forcing functions. Upon further review (sometimes after a mishap), it was determined that those test points were not achievable given known limitations. What practical things can be done to detect and alleviate those situations?
- Excessively Tight Test Point Tolerances
  - Across the decades, MANY millions of dollars have been wasted in pursuit of test points with excessively tight technical tolerances. This situation has been amplified when a given test point has excessively tight tolerances in multiple dimensions forcing the test pilot to accomplish nearly impossible feats of skill. In some cases safety buffers were unintentionally sacrificed in order to meet the technical requirements. The situation has been further exacerbated when there was an additional technical requirement to stay within those very tight, simultaneous tolerances for a long period of time. Sometimes after MANY repeats, the test pilot was able to find the right technique to meet all requirements. Another typical outcome has been for all concerned to eventually recognize that the program could not afford continued attempts at the intended test point(s) and the tolerances were relaxed to be more reasonable. Are the tight tolerances truly a requirement, or just someone's strong desire? What can be done to recognize and alleviate these situations BEFORE all those millions have been spent?

REFERENCES

1. MIL-STD-1530A, *Military Standard; Military Structural Integrity Program, Airplane Requirements*, 01 September 1972.
2. MIL-STD-1530C, *Department of Defense Standard Practice: Aircraft Structural Integrity Program*, 01 November 2005.
3. Office of the Chief Engineer, *Configuration for Improved Flying Qualities Testing*, AFFTC-TIH-91-01, Edwards AFB, California, July 1991.
4. Edward M. Kraft, The Air Force Digital Thread/Digital Twin - Life Cycle Integration and Use of Computational and Experimental Knowledge, 54th AIAA Aerospace Sciences Meeting, AIAA SciTech Forum, (AIAA 2016-0897)
5. Workshop Report, *Owning the Technical Baseline for Acquisition Programs in the U.S. Air Force,* National Academies Press, Washington D.C., ISBN 978-0-309-37431-6
6. MIL-F-87242, *Military Specification: Flight Control System General Specification For*, 30 August 1991.
7. MIL-F-9490D, *Military Specification: Flight Control Systems – Design Installation and Test of Piloted Aircraft General Specification For*, 05 September 2007.
8. JSSG-2008A, *Department of Defense Joint Service Specification Guide: Vehicle Control and Management System*, 07August 2014.
9. Damman, Lawrence M., et. al., *Flight Test Development and Evaluation of a Multimode Digital Flight Control System Implemented in an A-7D (DIGITAC) Volume 1,* AFFTC-TR-76-15-Vol-1, Edwards AFB, California, 01 June 1976.
10. Eggers, James A.; Bryant, Jr, William F., *Flying Qualities Evaluation of the YF-16 Prototype Lightweight Fighter,* AFFTC-TR-75-15, Edwards AFB, California, 01 July 1976.
11. Stoliker, Fred, et. al, *NASA F-8 Digital Fly-By-Wire Rate Limited Oscillation - Flight Testing at Edwards*, Flight Test Engineers Stories 1946-1975, Paperback-1996.
12. McLaughlin, Michael J.; Roth, Dan A., *YF-22A Engineering and Manufacturing Development Flight Test Summary*, AFFTC-TR-92-18, Edwards AFB, California, 01 October 1992.
13. Keller, Keith L., Newton, Robert M., *F-16C Block 40 Flying Qualities and Revised Pitch Integrator Rate Limits Flight Test Evaluation,* AFFTC-TR-96-24, Edwards AFB, California, January 1998.
14. Glaser, Scott T.; et. al., *F-22A Jet Wake Encounter Demonstration,* AFFTC-TR-06-16, Edwards AFB, California, 01 June 2006.
15. Charles H. Thornton and Perry Lamy, *Measuring Flight Test Progress on Large Scale Development Programs,* AIAA-92-4070-CP, American Institute of Aeronautics and Astronautics, Inc. with permission, 1992.
16. *Incorporating Test and Evaluation into Department of Defense Acquisition Contracts*, Office of the Deputy Assistant Secretary of Defense for Development Test and Evaluation, Cleared for Open Publication 24 October 2011, DOD 12-S-0150, October 2011.

APPENDIX A – LESSONS LEARNED IN BRIEFING SLIDE FORMAT

This page was intentionally left blank.

Envelope Expansion Lessons Learned Handbook

Paul Sorokowski
Michael Garland

Slides based on two workshop presentations:

- Edwards AFB on 8 August 2016
- NASA Armstrong Flight Research Center on 28 March 2017

Some modifications based on feedback during and after workshops

“Those who fail to learn from history are doomed to repeat it”

Winston Churchill (and others)

Long version:

“When the situation was manageable it was neglected, and now that it is thoroughly out of hand we apply too late the remedies which then might have effected a cure…

…Want of foresight, unwillingness to act when action would be simple and effective, lack of clear thinking, confusion of counsel until the emergency comes, until self-preservation strikes its jarring gong–these are the features which constitute the endless repetition of history.”

Excerpts from Churchill’s speech

to House of Commons, 2 May 1935

But… just like Cassandra of Troy:

“Enjoy this gift of prophecy…

…along with the curse that no one will believe you!”

A Simple Customer Question Motivated this Handbook

"What lessons have been learned during past envelope expansion flight test projects at Edwards Air Force Base"?

Envelope Expansion Lessons Learned

- People at EAFB have a HUGE amount of relevant experience
- Difficult to find a useful compilation of generally applicable lessons
- Envelope expansion projects will be fewer and further between
- Something was needed before lessons lost
- Started by talking to people…

People Interviewed to Mine Experience & Identify Topics

John Manke
Robert Lee
Kathy Wood
Mark Crawford
Tim Cacanindin
Brian Hobbs
Jessica Peterson
Chris Eaton
Fred Webster
Jason Bostjancic
Kirk Harwood
Reagan Woolf
Wendy Hashii

Intended Users

- Primarily 412 TENG Flight Test Engineers & Management
  - Mainly discipline engineers, multi-discipline leads
  - Engineers will need help from Chief Engineers and Deputy Directors

- May be helpful for other flight test organizations

- May not be appropriate for all customers or contractors
  - Unlikely to have patience for this scope of data dump
  - Probably better to discuss individual lessons as relevant

Handbook does NOT Cover…

- Lessons unique to any one project or organization
- Mission Systems or Sustainment efforts
    - Some lessons learned are general enough that they may apply to all disciplines
- The perspective of test pilots, test conductors, or maintainers

Handbook is NOT About…

- Developing flight test plans or safety packages
- Making flight test cards
- Executing Flutter, Loads, or Flying Qualities flight testing
- Analyzing flight test data
- Writing flight test reports

So What the Heck IS Handbook About?

- Identify recurring problems from many major flight test projects
  - Suggest methods to try to avoid those problems

- Identify methods that were successfully used to enhance test team preparedness and capability

- Primarily based on past projects for new production aircraft
  - Also research projects for cutting edge technologies (X-29, etc.)

- Most applicable to “new” aircraft flight test projects

How to Use Handbook

- Not a "Cookbook"

- Ponder the topics, then decide how to best apply to YOUR project
  (if at all)

- Interpret topics as:

  "Things that have gotten in the way of Envelope Expansion"

  Or:

  "Things that have helped enhance Envelope Expansion"

- Individual lessons may be useful during topic-related negotiations with program office or contractor

Emphasis on Capturing Experience

- No attempt to "prove" one method versus another
- No comprehensive, useable database
- Most prior customers would not have paid for studies to obtain data to document bad decisions
  - Could supporting data be obtained WHILE projects are underway? (maybe… good luck with that)

Disclaimer

The opinions expressed during this presentation and in the handbook do not necessarily represent those of management or anyone other than the authors

Presentation Format

- Get familiar with contents of Handbook
  - Slides are just "Tip of the Iceberg"
  - Read Handbook later

- 22 topics
- 2-4 slides per topic
- Lots of words… no pictures… just deal with it
- Interactive discussion
  - Try to wait until after slides on that topic
- Target 15 minutes per topic
- Send us comments not discussed (see last slide)

Two Major Sections

- TOPICS TO COORDINATE PRIOR TO CONTRACT AWARD
  - Implies "Early Tester Involvement"
  - Trickier if testers don't get involved until after contract award… but still worth pursuing
- TOPICS TO COORDINATE PRIOR TO FIRST FLIGHT
  - May still require ongoing negotiations after first flight
- No section for "TOPICS TO COORDINATE AFTER FIRST FLIGHT"
  - Probably too late… but never give up!

TOPICS TO COORDINATE PRIOR TO CONTRACT AWARD

- CA1: Don’t Neglect Basic Aircraft Systems
- CA2: EMD Test Aircraft Configuration & Utilization
- CA3: Expected Test Efficiency for Test Program Planning
- CA4: Implement Low Cost Flight Test Support Simulator
- CA5: Incorporate Flight Test Aids on Test Aircraft
- CA6: Plan for Adequate Number of Software Revisions
- CA7: Plan for Shifting Work-Split within Test Team
- CA8: Tracking Flight Test Progress
- CA9: Update System Models with Flight Test Data
- CA10: Flight Test Fatigue Failures Due to Flight Test Exposure Beyond Operational Design Usage
- CA11: Government Access to Fundamental Aircraft Models

TOPICS TO COORDINATE PRIOR TO FIRST FLIGHT

- FF1: Confidence Testing Prior to Delivery of New Capabilities
- FF2: Flight Test Flexibility
- FF3: In-Flight Simulation Prior to First Flight
- FF4: Multi-Discipline Envelope Expansion
- FF5: Providing Focus on Fixing System Deficiencies Impacting Flight Test
- FF6: Prudent Focus of Flight Test Points
- FF7: Qualitative Evaluations
- FF8: Residual Oscillations
- FF9: Rate Limited Control Surface Oscillations
- FF10: Plan for Independent 3rd Generation Data Analysis Tools
- FF11: Crosswind Takeoff and Landing Envelope Clearance

TOPICS TO COORDINATE PRIOR TO CONTRACT AWARD

CA1: Don't Neglect Basic Aircraft Systems

- Do not delay implementation of basic systems as a tradeoff to obtain an "early look" at more advanced weapon system technologies
  - Basic systems are cornerstone of weapon system integrity & operability
  - Considered "routine" & low risk, often neglected for sake of advanced cutting edge technologies
  - Profound consequences to efficient progress of test program and weapon system development

- Focus early testing on basic systems & expanding flight envelope
  - First few test aircraft configured with appropriate basic systems, structure, instrumentation, and flight test aids to enable focus on basic systems
  - Minimize/eliminate use of "off-the-shelf" systems in lieu of the intended system
  - Have contingency plans in place to deal with inevitable problems

- Encourage early mission systems evals, but not at expense of basic systems
  - Find the right balance!

CA1: Don't Neglect Basic Aircraft Systems

- Examples:
  - Communication, navigation, weapon bay doors, structure, brakes, landing gear, environmental control systems, fuel system, hydraulics, etc.

- Typical planning decisions that can be counterproductive
  - Non-production representative, "off the shelf" systems
    - Only used on 1st or 2nd test aircraft
    - Highly disruptive when they don't work right, lots of wasted effort to fix
  - Assumption that system is "mature" due to prior application
    - Little or no development or preparation
    - Integration issues VERY underestimated

- Probably needs to be addressed well above discipline engineer level

CA2: EMD Test Aircraft Configuration & Utilization

- 1st test aircraft under-utilized & unsuitably instrumented
  - Contractor incentivized to meet delivery date
  - Design schedule slips, capabilities sacrificed to meet delivery date
    - Not production representative with significant unplanned flight limitations
    - Not configured or instrumented for other test types
    - Other test types could be conducted even within flight limitations

- When planning EMD test aircraft configuration & utilization:
  - Incentivize contractor to deliver critical capabilities, not just aircraft delivery date
  - Planned configuration should reflect inevitable design delays & unplanned limitations
  - Avoid narrowly focused purposes for first few test aircraft
  - Plan instead for broader general utility
  - Much mission systems work can be accomplished in limited envelope
  - Provide early start on very long, critical paths

CA2: EMD Test Aircraft Configuration & Utilization: Example

- Impact of typical structural design process
  - EXTREME emphasis on weight reduction to meet performance requirements
  - Leads to optimistic interpretation of structural margins
  - Later, more rigorous analyses reveal dozens or hundreds of negative margins
  - Many negative margins not fixable prior to first flight
  - Each negative margin requires some type of flight limitation
  - Some problems fixable with retrofits and lengthy downtime
  - Many problems can only be fixed on later aircraft
  - End result is permanent limitations on first aircraft
  - First aircraft planned and instrumented as a structural test "workhorse"
  - Much of the instrumentation and work must be shifted to a later aircraft
  - Ship 2 often too close behind Ship 1 to implement fixes
  - Sometimes Ship 3 can be built with production representative structure
  - Ship 1 retires early as: maintainer trainer, live fire target, decoration, or museum piece

CA2: EMD Test Aircraft Configuration & Utilization: Possibilities

- Instrument Ship 1 to clear a useful but partial envelope
  - Most likely end result anyhow
  - With less structural instrumentation… get to first flight sooner!
  - Must not allow any design that places partial envelope at risk
    - Make sure there is LOTS of margin
- Task Ship 1 with “easy” flight tests
  - Things that can be done in partial envelope
  - Aerial refueling, propulsion, subsystems, etc.
  - Mission systems, initial weapon bay door work
- Designate later aircraft (#3?) for full up structures instrumentation
  - Allow time for structural design to stabilize
  - Consistent with MIL-STD-1530

CA3: Expected Test Efficiency for Test Program Planning

- Planning typically uses flight test efficiency factors greater than statistical efficiencies from previous similar programs
  - Actual flight test efficiencies much less than planned
  - Leads to extended test schedules, resultant higher costs, and appearance of poor program progress
  - Number of test point repeats needed also underestimated

- Flight test efficiencies historically average 60% to 65% over the course of the program
  - 60 to 65% of the flying time is spent in the actual execution of test points
  - Test point repeats historically 40% to 60% of planned test points

CA4: Implement Low Cost Flight Test Support Simulator

- Locate at test site
  - No air travel required, minimal commute by auto ok
- Hardware-in-loop or Iron Bird NOT required
- Sufficient with software models only… but hi-fidelity
  - Best models provided by contractors
    - Model support for test team part of contract
    - Use models that already exist… not special for flight test
    - Must stay current with flight test configurations
  - Primary simulation models can be executable code
    - Best results when aerodynamic and control law models are source code (written into contract)

CA4: Implement Low Cost Flight Test Support Simulator

- Rudimentary cockpit ok
  - Primary control devices representative
    - Stick/Yoke & Throttle with same feel and characteristics as test aircraft
  - Other PVI items can be “emulated” with touch screens, etc.
  - Low cost “out the window” display setup

- Dedicated to flight test support
  - Must be available whenever needed by flight test team
  - Can not split scheduling with other development tasks

CA4: Implement Low Cost Flight Test Support Simulator

- AVOID simulator at remote site as only option
- Test site sim can be operated by contractor or USAF
  - Should be much cheaper than primary development simulators
  - Need mini-control room at test site sim, ideally with same displays as MCR
  - Connecting to main control room ok for early team mission rehearsals
    - Expensive, hard to schedule
- Give consideration to Emergency Procedures trainer capability
  - If good enough for pilot qualification, can save TDYs
  - Potentially very valuable for test team (if representative of test aircraft)
  - Can increase costs
- Beware of delivery “just like the operational trainer”
  - Insist on flight test capabilities
  - Data recording & displays, flight test aids, off-standard day temps, etc.

CA5: Incorporate Flight Test Aids on Test Aircraft

- Many types of Flight Test Aids (FTAs)
  - Dial-a-g, dial-a-gain, flutter excitation, preprogrammed test input (PTI), etc.
  - Special indicators for pilot on HUD or other display
  - Also needed for unique maneuvers and test techniques difficult to perform manually
- Contracts often neglect to specify requirement for FTAs
  - Lack of FTAs would lead to inefficient testing
    - More repeats, more potential for exceeding interim limits, wait on OFP updates to evaluate fixes
  - No clear contract requirement causes unnecessary distractions
    - FTAs must be justified through laborious negotiations (FTAs eventually judged necessary)
    - Must be funded by alternate "out-of-scope" mechanisms
    - FTAs and pilot interface implemented as an "afterthought", not well integrated with design
- Experience shows FTAs provide test efficiency improvements & limited capability to respond to unpredicted results without waiting for OFP update

CA5: Incorporate Flight Test Aids on Test Aircraft

- Get generic requirement for FTAs into contract language
  - Way too early to define specific types of FTAs
  - Goal: make it very clear that FTAs are "in-scope"
- If too late to influence contract:
  - Assemble generic examples from previous programs
  - Don't use specific implementations that prior contractors may consider proprietary
  - Justify FTAs by obtaining consensus within discipline level at program office and contractor
    - Also need support from Test leadership
- Once FTAs authorized (via contract or other negotiations):
  - Work with discipline counterparts to define specific implementations
- Enjoy your more efficient, effective, and safe test project
  - Pass lessons learned on down the line

CA6: Plan for Adequate Number of Software Revisions

- Number of planned software revisions typically WAY underestimated
  - #1 to get started, #2 to add capabilities and a few fixes, #3 to complete capabilities and final fixes
  - Actual number of revisions typically 2 to 10 times higher (or more)
  - How does all that extra workload get funded? (design work, V&V work in labs, retest)
  - Visibly results in extended schedules or de-scoped testing with added development risk

- Systems with significant advancements in technology need more revisions than evolutions of existing systems/capabilities
  - Maturity of the technology is overestimated
  - Degree of difficulty to advance maturity is very underestimated

- A robust and honest assessment of the maturity of the technologies may enable more accurate projections of number of revisions
  - Legacy programs for new aircraft provide a sense for number of revisions required vs planned

CA6: Plan for Adequate Number of Software Revisions

- Classic quotes:
  - “If we run into minor problems we’ll just use workarounds or do something else until the next major software update”
  - “If we run into a show-stopper problem we can pump out an emergency update in less than 2 weeks”

- Planned major updates can be 12-18 months apart
  - Cycle time between problem ID in test and FIRST attempt at a fix can be longer
    - Need time to analyze problem & design fix
    - Missing software freeze date bumps fix to FOLLOWING major update
    - OR… major update with new capability must be delayed
  - When a fix misses a major update, interim limits and workarounds need to remain in place

- Software revision plan needs to be reasonable
  - Major updates with new capabilities @ once per year – Ok
  - Need PLANNED interim updates for flight test fixes
    - Every 3 months possibly sufficient… but need to find the right “rhythm”
    - Once per month historically common… but probably excessive as “The Plan”

CA6: Plan for Adequate Number of Software Revisions

- Prior to contract award:
  - May be very difficult to define well-targeted contractual language
  - Advocate multiple system integration test lines for software/hardware V&V flow (allows for new development simultaneous with flight test fixes)
  - Advise program office about unreasonable assumptions in contractor proposals
  - Not the job of flight testers to get contractor to change proposal

- After contract award:
  - No easy solution
  - Attempt to negotiate a reasonable software revision plan
    - Way above discipline level, but may need discipline help
    - Initial reaction probably underwhelming
  - Be persistent and try different approaches
  - Ultimately, reality WILL prevail
  - Oh yeah… develop metric to track cycle time between flight test problem ID and the FINAL fix

CA7: Plan for Shifting Work Split Within Test Team

- Contractors almost always plan to support an adequate, long term presence at test site
  - Classic quote: "We HAD to plan our staffing as if there was zero USAF support"

- Adequate long term contractor presence almost never happens
  - Initial cadre of high-caliber talent called back to home office
  - Can't convince other home office engineers to uproot long term to test site
    - Moving expenses and per diem zeroed out due to ongoing budget problems
  - Difficult to hire new engineers willing to live near test site
  - Two week rotations from home office become common

- Results:
  - USAF engineers pick up more of the workload to "turn the crank"
  - Designers lose first hand insight into problems encountered, delaying resolution of issues

CA7: Plan for Shifting Work Split Within Test Team

- Prior to contract award:
  - Advocate stable contractor and government staffing throughout project
  - Recognize historical trend of shifting work-split as part of the plan
    - Somewhat biased towards contractor support initially
    - Fairly rapid shift to 50-50
    - Heavily staffed by USAF in last half
  - Initial program office reaction likely to be disbelief
    - May be helpful to dig up historical trend data
  - Unlikely that contract language will be influenced
  - At least the concept may begin to gain traction

CA7: Plan for Shifting Work Split Within Test Team

- After contract award:
  - Acknowledge and plan for heavy contractor staffing initially
  - First flight mission control room can be near 100% contractor
    - It’s their baby, big prestige in contractor world
    - No need to mandate USAF support, only as requested or needed
  - Mission control room for early flights can be mostly contractor
    - If three engineers needed for a discipline, plan 2 for contractor
    - Enables USAF engineers to gain knowledge from experts
  - Once into a flight test rhythm, 50-50 staffing is a GOOD plan
    - Provides for a well integrated team
    - Enables high-caliber contractor talent to return focus to ongoing development at home
    - Enables smooth shift over time to USAF support
  - Avoid a USAF staffing “panic” when the inevitable trend takes its toll on contractor staffing

CA8: Tracking Flight Test Progress

- Traditional flight test progress measurement techniques not adequate for large scale development programs
  - Primary example: Test points completed and overall flight test hours
    - Easy to track but misleading
    - Excessively simple metrics have driven bad decisions
  - Lack true measure of progress and work to go for adequate program control
    - Not all test points are equal (some take MUCH longer)
- Techniques used on some programs provided a more realistic measure of progress toward program milestones
  - Measured work scheduled, work completed and work remaining
  - Normalized for more accurate measurement of work
  - More integrated flight test planning
  - Insightful flight test progress measurement
  - Enabled more effective flight test program control
  - Provided a more credible and meaningful report of flight test status
    - Program Office
    - Service Department
    - DOD Acquisition Manager

CA8: Tracking Flight Test Progress

- Examples of past bad behavior due to poorly conceived metrics:
  - Test point "burn down" charts dominate all decisions
    - Test points needing a minute or less scored the same as test points that required multiple flights
    - To provide an artificial appearance of progress, relatively easy test points get accomplished early
    - Many difficult test points NEED to be attempted early
    - Building bow wave of work not predicted/visible (repeats, regression, longer test points)
      - Near end of EMD, all remaining test points very difficult or require multiple flights
  - Excessively simple metrics can contribute to disruptive and unproductive mandates from external entities
    - Such as: "Complete 5% of total flight test hours before more funding provided"
    - Result on more than one project: Flying solely to build flight time with no productive testing
    - Mandates more meaningful if based on capabilities tested
      - Can still drive bad behavior

CA8: Tracking Flight Test Progress

- No easy solution
- Examples of appropriate actions:
  - One method of normalization based on “time per test point”
    - Need realistic estimates, stakeholders motivated to minimize their “pile of beans”
    - Flight testers need to be the neutral and fair “cops”
  - Develop appropriate bookkeeping
    - Overhead time for takeoff, aerial refueling, RTB not counted to individual test points
    - Setup time to achieve special test conditions should count to individual test points
      - Fuel transfer for aft CG, climb to high altitude, fly to remote range
  - USE CAPABILITY-BASED METRICS!
    - Built around capabilities based objectives and milestones
    - More important than fundamental metric “beans”
    - Keep focus on operational requirements
    - Flight testers can advocate, but needs “buy-in” at many levels

CA9: Update System Models with Flight Test Data

- Flight test data has not been used to provide comprehensive updates to the system models as needed to support ongoing development and sustainment
  - Models generated from sources that provide a theoretical representation
    - Produces predictions with various levels of credibility
  - The original (or partially updated) system models continue to be used for:
    - Resolving system development issues, providing flight test predictions, providing system performance data in flight manuals for aircrew use, and for subsequent system developments
- If the system models were comprehensively updated with flight test data:
  - Subsequent predictions would be more accurate, more reliable, reducing development risk
  - Aircrews would have more accurate performance data
  - Subsequent flight test scope and risk would be reduced given more accurate predictions
  - Possibly preclude the need for repeat flight tests later in the system life cycle
- Significantly leverages the money spent for flight test

CA9: Update System Models with Flight Test Data

- Classic quotes (danger signs):
  - "Of course we'll update the model, but only if the differences are big"
  - "This sounds like a science project to me"
- Perceptions of the importance for model updates varies at all levels
- Even when initial intentions are good, budget cuts and priority issues seriously degrade the product actually delivered
- Recent acquisition initiatives may be helpful
  - "Digital Thread"
  - "Own the Technical Baseline"
- Related to CA11: Government Access to Fundamental Aircraft Models

CA9: Update System Models with Flight Test Data

- Appropriate Actions:
  - Find the right balance between accurate updated models and a “science project”
  - Attempt to include contractual language to require regular updating of models
    - Specify a reasonable model update rate to reflect flight test results
      - “Planned updates every 12 months” may be workable, every 2 months seems too short
      - Include additional clause for “As Needed” to support ongoing flight test
  - Regardless of contractual language, develop team philosophy for model updates
    - Early discipline-level discussions with program office & contractor
    - Need team philosophy for model updates even if government denied access to models
  - Need clear ground rules for utilization of updated models
    - Be wary of highly segmented model updates with fuzzy boundaries of applicability
    - Ideally, a single “certified” model automatically compensates for odd “cut out” regions
    - Inability to define boundaries of applicability may signal need for more flight test data

CA10: Flight Test Fatigue Failures Due to Flight Test Exposure Beyond Operational Design Usage

- Test programs typically do not plan for the increased exposure required to flight test for full envelope clearance
  - Cumulative impacts of multi-discipline testing (flutter, loads, FQ, propulsion, vibroacoustics, etc.)
  - Envelope expansion testing requires significant time at high dynamic pressure (or other critical environmental conditions) to clear envelope
  - Conversely, design operational usage segments typically allow for very little time in critical environmental conditions (based on planned operational profiles)
  - Omission has led to in-flight failures of fatigue critical components during flight test

CA10: Flight Test Fatigue Failures Due to Flight Test Exposure Beyond Operational Design Usage

- Examples:
  - Cracked Fuselage Panels on Legacy Twin-Engine Tactical Fighter
  - Pressure Relief Valve Failure on High Performance Fighter
  - Configurable Rail Launcher (CRL) Arm Cracks on High Performance Fighter
  - Trailing Edge Flaperon Failures on a Legacy Single Engine, Multi-Role Fighter

Flight Envelope as a Function of % DLL on the Flaperon	Hours / 8000 Hour Life	% of Operational Flight Time Spent in Envelope
<60%	7680	96%
<85%	312	3.9%
<100%	8	0.1%

CA10: Flight Test Fatigue Failures Due to Flight Test Exposure Beyond Operational Design Usage

- Appropriate Actions:
  - Prior to initiation of flight test, identify components at risk of failure during flight test
  - Track the flight test usage of at-risk components (develop methodology prior to initiation of flight test)
  - Develop a plan for how to handle fatigue critical components during flight test (increased inspections, replacement parts on hand at the CTF, etc.)

- Early identification of at risk components can save time/money due to down time caused by unexpected test events, waiting for replacement parts, needing to develop fixes, etc.

- Also… consider adding words in Flight Manual to inform pilots about repeated and prolonged exposure near edges of flight envelope

CA11: Government Access to Fundamental Aircraft Models

- All flight testers must have deep understanding of system under test
  - Best obtained with access to fundamental models
    - Control laws, aero model, engine model, etc.
  - Historically VERY difficult
    - Contractual constraints & lack of understanding
  - Recent trends provide opportunities for improvement
    - Early tester involvement (to influence contractual documents)
    - Acquisition initiatives: ”Digital Thread”, “Own the Technical Baseline” (to provide “top cover”)
- Future contractual language may have useful “concept definitions”, but considerable ambiguity may remain
  - Test teams may still need to negotiate to obtain the most workable solution
  - Contractor will still be highly motivated to protect “intellectual property”
- Best practices: Goals for testers during early program planning
  - Explain WHY testers need access to models
  - Advocate explicit contractual language or other post-contract agreements
  - Accept compromises as needed to obtain most workable solution

CA11: Government Access to Fundamental Aircraft Models

- Concepts that may help improve flight tester understanding
  - If model is stamped "Proprietary" or "intellectual property" doesn't mean that it is true
    - Many contractors tend to put those stamps on anything and everything
    - Government agencies have often been weak in challenging those assertions
  - Examples of items that are not "Proprietary" or "intellectual property":
    - Models obtained from government agencies
    - Models from the public domain (e.g. generic atmospheric models)
    - Models developed using taxpayer dollars in direct support of a project (depending on contractual language)
    - Ideally, the Program Office should be the adjudicator… but often needs help from test management
  - The term "Proprietary" does not necessarily mean that the government cannot have access to models (Intellectual Property probably has more constraints)
    - If the models are truly Proprietary, the content needs to be protected accordingly
    - Existing non-disclosure agreements may be sufficient
    - If existing contractual language or other agreements are not adequate, a special agreement within the working group may suffice (see next slide)

CA11: Government Access to Fundamental Aircraft Models

- Examples of working group agreements and compromises that have been successful
  - Accept executable code if obtaining source code becomes stuck in legal battles
    - Having access to a useable model is better than nothing
  - Agree to coordinate before elevating any issues discovered
    - Provide an opportunity to develop a consolidated team position
  - Don’t ask for anything that doesn’t already exist
    - Added features = added workload/cost
    - May be able to convince technical people that they need proposed feature for their purpose
  - Don’t ask for internal analysis tools that have been developed at contractor expense
- Persistence pays off… in either direction
  - Don’t give up after first rejection
  - Work with a different person or organization to obtain support
  - Try a different tactic or justification
  - Use your knowledge of the individual motivations
  - Contractors can be very reluctant, but can sometimes envision the benefits of sharing

TOPICS TO COORDINATE PRIOR TO FIRST FLIGHT

FF1: Confidence Testing Prior to Delivery of New Capabilities

- Historical Problem:

  New capabilities delivered but not actually ready for flight testing

- Main Goal: Expose new capabilities to realistic conditions and procedures that will soon be experienced during flight tests (versus canned V&V runs)
  - Provides a very effective check & balance to assess readiness prior to expensive flight testing
  - Important Bonus: Test team gets improved understanding of new capabilities and better preparation

- Frequently skipped to reduce short term costs and schedule impacts
  - New products delivered not ready for test or with serious limitations to test
  - Rush to meet delivery date often overlooks question:

    "Can the product execute as needed over the next few months of flight test"

FF1: Confidence Testing Prior to Delivery of New Capabilities

- Appropriate Actions:
  - Test team should champion concept long before flight testing begins
  - Need to obtain support from program office and contractor management
    - Can be convinced that it is in their best interests
  - May need to combine with existing planned "graduation exercises"
- Fundamental Requirements:
  - Piloted simulator with most up to date aircraft system software versions hosted on the most relevant aircraft computer hardware with hi-fidelity models for other components
  - NOT iron bird (prohibitive expense, scheduling, and difficulty of operation)
  - Small, representative flight test team conducts evaluation
    - Typical minimum: Test pilot, key discipline engineer(s), test conductor
  - Ideal time is soon after normal lab V&V (typically 1-2 weeks prior to flight tests)
    - Sooner runs risk that changes implemented after test could introduce undiscovered issues
    - Later risks insufficient time to react to unexpected results

FF1: Confidence Testing Prior to Delivery of New Capabilities

➢ Confidence Test Do's and Don'ts

- DO NOT get hung up on the term "Confidence Testing"
  - Use whatever name works for the team
- DO NOT allocate excessive time for Confidence Testing
  - Minor mods = a day or less
  - Major new capabilities = 2-3 days
- DO NOT repeat requirements-based testing from Verification & Validation
  - Focus on key upcoming flight tests
- DO NOT confuse Confidence Testing with normal test team mission rehearsals at test site
- DO accomplish Confidence Testing in the most representative ground-based simulator available
- DO require that Confidence Testing be accomplished by a small team from the flight test community (not the designers)
- DO attempt to provide simulator control room team with displays representative of those available at test site

FF2 – Flight Test Flexibility

- Projects often establish a single path thru multi-discipline test matrices
  - Envelope expansion defined as an explicit series of specific test points
  - Key stakeholders refuse to consider alternate paths
  - Any deviations must be re-coordinated at high levels
  - Weeks or months of delay when inevitable problems encountered
- Alternate flight test paths need to be enabled by using common-sense ground rules instead of a single, inflexible test point sequence
  - A “preferred” path can still be defined
  - Ground rules provide logical decision paths when preferred path derailed
  - Enables test team to proceed without lengthy coordination
- On previous projects, it was eventually possible to establish reasonable ground rules after the single path had been derailed more than once

FF2 – Flight Test Flexibility

- Flight test management must direct all envelope expansion stakeholders to coordinate and establish flexible ground rules
  - At any given point along the preferred envelope expansion path, common-sense ground rules must provide for at least one alternate path
  - Examples:
    - Instead of stepping to next higher Mach number at same altitude, go to lower Mach number at lower altitude
    - Instead of stepping to next lower altitude at same Mach number, go to lower Mach number at lower altitude
    - Allow test points from another discipline as long as “far enough” behind “leading edge”
- Establish flexibility principles long before disciplines establish test concepts
  - Advanced Planning Team, Long-Lead Team (or whatever) hammers out specifics
  - Once stakeholders accept flexibility principles, process may actually be easier than coordinating a single path

FF3: In-Flight Simulation Prior to 1st Flight

- Can be very helpful confidence builder
  - Key advantages over ground-based simulations (fixed or motion)
  - Focus on main uncertainties associated with 1st flight and initial envelope expansion
    - Primarily landing tasks
    - May be helpful for aerial refueling tasks, even without boom contact
  - Particularly important to reduce development risk for cutting edge designs
    - Innovative controllers
  - Significant control law changes implemented as a result of in-flight simulation
  - Improves confidence during First Flight Readiness Review process
- Test & design teams should seriously consider long before 1st flight
  - Genuine evaluation of pros and cons
  - Biggest constraint may be retirement of many aircraft from in-flight simulation fleet

FF3: In-Flight Simulation Prior to 1st Flight

- Preparation is essential for successful in-flight simulation
  - Commitment of data, engineering time, and management coordination
  - If treated as after-thought, results not worth the investment
- Tricky to make one aircraft fly like another
  - Example: May match pitch rate response, while sacrificing load factor response
  - Understand constraints; work closely with in-flight sim personnel to minimize impacts

FF4: Multi-Discipline Envelope Expansion

- Related to FF2: Flight Test Flexibility
- Envelope expansion initially planned as single-discipline stovepipes
  - Postpones inevitable recognition that flight test MUST be accomplished as integrated effort
  - Inefficient flight testing when discipline-selected test conditions do not match
    - Each discipline selects test conditions to match different design tools (wind tunnels, analytical models, etc.)
  - Aircraft not always appropriately instrumented for integrated testing
    - Individual discipline tests targeted for different tail numbers
  - Delays envelope clearance for other testing to begin
- Expediency for accounting purposes should not usurp integrated multi-discipline approach to envelope expansion testing

FF4: Multi-Discipline Envelope Expansion

- Very early in overall process, flight test management must mandate integrated multi-discipline test planning
  - Contractors often begin stovepipe planning well before contract award
  - Need to get each discipline working together towards similar goals before incompatible concepts become too "cemented"
  - Need to identify each discipline that needs to be part of integrated testing on a given test aircraft to insure compatible instrumentation
- Integrate envelope expansion planning and flight tests
  - Key envelope expansion disciplines need to collaborate on test conditions, test maneuvers, build-ups, and instrumentation to simultaneously satisfy each respective requirement
  - Consolidate multi-discipline test points using "integrated test blocks". Examples:
    - Leading Edge Integrated Test Block – First time test conditions, typically at or near 1g
    - Maneuvering Integrated Test Block – Follows behind LEITBs, often to expand load factor, rolling maneuvers, etc.
    - Trailing Integrated Test Block – After LEITBs and MITBs, completes envelope clearance for all disciplines

FF5: Providing Focus on Fixing System Deficiencies Impacting Flight Test

- Related to CA6: Plan for an Adequate Number of Software Revisions
- Many large EMD-type projects have followed a consistent pattern:
  - "Flight testing is only a verification step to confirm that the system meets requirements… and maybe fix a few unexpected problems"
  - Given that philosophy: VERY few software updates planned, highly focused on new capabilities
  - During flight test: MANY problems on multiple systems, update process overwhelmed
  - Schedule for next planned software update already very tight, so addition of any flight test fixes would delay delivery of important capabilities (often tied to incentives)
  - Many flight test problems not considered "emergencies", unaddressed for 1-2 years or longer
  - Eventually, a reasonably aggressive process to provide fixes was developed
    - Sometimes resulting in software releases every few weeks (probably too often)

FF5: Providing Focus on Fixing System Deficiencies Impacting Flight Test

- But… software update management “graded” on how many internal problem reports were “burned down”
  - Problems selected for fixes were relatively easy to burn down (regardless of impact to ongoing flight test)
  - Internal problem reports CLOSED as soon as fix designed or included in an upcoming software release
  - When the original problem was not fixed, a NEW internal problem report opened (cycle often repeated)
- Sense of progress portrayed by typical “red light-green light” charts was actually far less than actual context of progress
- Resolution of major issues affecting operational capability and testability delayed
- During that entire pattern, the total duration between flight test ID of a problem and the FINAL fix, was typically not tracked as a meaningful programmatic metric

FF5: Providing Focus on Fixing System Deficiencies Impacting Flight Test

- Flight test management needs to aggressively advocate that fixes for flight test problems be built into overall plan, not just an afterthought
  - Include enough capacity in labs and V & V process to handle simultaneous delivery of fixes to flight test problems along with development of new capabilities
  - In between major updates for new capabilities, plan on regular updates dedicated to flight test fixes
    - A reasonable update rate might be every 3-4 months
  - Seriously consider a programmatic metric to track the total duration between flight test ID of a problem and the final fix… using a priority scheme as described on next slide

FF5: Providing Focus on Fixing System Deficiencies Impacting Flight Test

- Incentivizing and tracking the burn down of flight test problems needs a sense of relative magnitude applied to each problem
  - Focus on most serious problems with the biggest impacts on operational capability and impact to efficient progress of flight test
  - Higher credit for resolution of a serious problem, lower credit for a simpler problem with lesser consequences
  - Could be as simple as a priority scheme, or a weighting of each problem
  - WITs and DRs need to provide description of impact of problem on operational capability and also impact on flight test
    - But… don't rely exclusively on DR system to get flight test problems fixed… things get "lost"
    - Test team needs another way to maintain focus on flight test issues

FF6: Prudent Focus of Flight Test Points

- Excessive flight testing tends to be conducted in the "heart of the envelope"
  - Many projects have used a nice symmetric matrix with equally spaced test points
- Typical trend contributes to later pressures to reduce test scope just when the more critical envelope edges are approached
- Areas where predictions typically good warrant less scrutiny
  - Take larger steps, reduce scope of testing
- Areas where predictions most suspect warrant more scrutiny
  - Near or at the edges of the envelopes
  - Anywhere predicted performance is suspect (e.g. transonic)
  - Use finer increments, more configurations (for safer buildup, better coverage)
  - However, recognize each discipline may define "heart of the envelope" differently
    - For example, the "edges" of the Loads envelope (i.e. due to higher control surface deflections) may be encountered in the "middle" of the Flying Qualities envelope

FF6: Prudent Focus of Flight Test Points

- Prudent test point focus relies on extensive use of high fidelity models and simulation
  - Use simulation predictions to compare with flight test
  - Use predetermined continuation criteria to define bounds of agreement and course of action when flight test results do not match predictions
  - Best results if models updated regularly to reflect flight test results

- Best practice
  - Less time spent in heart of envelope may help diminish pressures to reduce scope as edges of envelope are approached
  - Regardless of flight test scope pressures, thinner testing in heart of envelope still prudent when high fidelity models and simulations available

FF7: Qualitative Evaluations

- Flight test will always be necessary to weed out characteristics objectionable to pilots
  - Quantitative predictions/tests may not manifest poor flight or system characteristics
  - Simulation useful for initial design, but limited relative to real flight environment

- Subject to individual pilot habits/preferences/experiences/training
  - Requires reasonable number of repeats by different pilots
    - Traditionally 3 to 6 pilots
  - Overall result heavily influenced by statistical relevance
  - Need balance between expense of achieving high statistical relevance and lack of statistical relevance

FF7: Qualitative Evaluations

- Trend on recent programs to cut flight test qualitative evals to spot checks
  - 1 or 2 pilots
  - Requirement for multiple pilot evals perceived as standing in the way of test point "burn down"
  - Cost cutting without understanding of potential consequences
  - Likely that significant issues discovered later with greater expense to resolve

- Testers work with contractors and program office in early planning
  - Establish understanding and philosophy for flight test qualitative evaluations
  - Search for a reasonable balance
    - Don't need 3 or more pilots for EVERY qualitative test point
    - "Build-up" evals may be sufficient with 1 pilot
    - Save evaluations with 3+ pilots for test points targeted at fundamental aircraft capabilities

FF8: Residual Oscillations

- Continuous aircraft oscillations in any axis not sustained by pilot inputs
  - Most common flight control/flying qualities problem first discovered in flight test
  - Many types: Limit Cycle Oscillations (LCOs), Structural Resonance, low damping from low gain or phase margins (NOT PIOs, NOT ASE)
  - Unpredicted system oscillations a sign of basic instability
- Hard to predict every contribution
  - Aircraft models used in simulations prior to flight test not fully representative
    - Aerodynamics, mechanical non-linearities, other real-world characteristics
- Test teams often not prepared
  - Likelihood of occurrence in flight test not well understood
  - Pre-defined criteria for continuing or stopping envelope expansion not established
    - Continuing without criteria could result in mishap
    - Stopping without criteria could result in needless test delays

FF8: Residual Oscillations

- Best practices
  - Gov team members participate in ground testing on test aircraft and/or iron bird
    - Limit Cycle Testing (not always in baseline contractor plan)
      - Simulate aerodynamic loop closure
      - Show that key surface effectiveness terms can be doubled (+6 db) without oscillations
      - On test aircraft in olden days, nowadays on iron bird
    - Structural Mode Interaction - SMI Testing (aka Structural Resonance or Ground Resonance Testing)
      - Always on test aircraft
      - Focused on structural feedback paths, usually without aerodynamics (sometimes combined)
      - Show control law filters can handle key loop gains of +6 db or higher without oscillations
    - Flight test team benefits from understanding: test conditions & procedures, impact of results on flight test (i.e. which flight conditions result in lowest stability margins)
  - Coordinate & pre-define continuation criteria if oscillations encountered
    - Example: Amplitude of oscillation of specific key parameters

FF9: Rate Limited Control Surface Oscillations

- Rate limiting part of every aircraft design since Wright brothers
  - But… major factor in the most insidious flight test incidents in last three decades
  - Especially on augmented flight control systems
  - More rate limited oscillations than predicted is a sign of serious stability problem
  - Specs/standards vague on boundary between excessive vs adequate rate limiting
- Predictability varies
  - Function of design philosophy aggressiveness regarding rate limiting
  - Models used in simulations prior to flight test not fully representative
    - Unusual flight conditions, unexpected trigger events, other non-linear phenomenon
- Test teams often not prepared
  - Likelihood of occurrence in flight test not well understood
  - Pre-defined criteria for continuing or stopping envelope expansion not established
    - Continuing without criteria could result in mishap
    - Stopping without criteria could result in needless test delays

FF9: Rate Limited Control Surface Oscillations

- Best practices
  - Aggressive design philosophy to detect and eliminate worst cases before flight test
  - Obtain team consensus on continuation criteria
  - Review time history predictions for each planned flight test point
    - To understand if rate limiting is predicted, and if so… to what degree
  - Actively look for sustained rate limiting during each test point
    - Do not proceed if pre-defined criteria exceeded (or aircraft response judged to be inappropriate)
  - Pilot and test conductor fully informed of mishap potential
    - Neither has any indication of excessive rate limiting… other than control room FQ engineer
    - Pilots tend to think they can get better control if they just try harder (which can make it worse)

FF10: Plan for Independent 3rd Generation Data Analysis Tools

- Enables TW confidence in officially published results
  - Provides veracity of substantiating data and analyses critical to credibility and withstanding scrutiny of results, C&Rs
  - In depth understanding of what's being done to the data
  - Focus of tools reflects flight test experience from similar systems
- Rough definitions:
  - 1st generation = basic conversion to engineering units from "raw" data
  - 2nd generation = standard fundamental calculations

    (air data & CG corrections, time alignment, smoothing, etc.)
  - 3rd generation (contractor) = spec compliance, Airworthiness assessment, requirements verification, model verification, verification of design considerations
  - 3rd generation (government) = multi-flight summaries, qualitative assessments, capability assessments, Military Utility, independent Airworthiness assessment, independent requirements verification, readiness for OT, etc.

FF10: Plan for Independent 3rd Generation Data Analysis Tools

- Sometimes programs drive use of common set of tools as cost savings
  - Usually developed by contractor with contractor focus
  - USAF testers not given access to code for scrutiny (intellectual property)
  - Attempts to include USAF tester's analysis needs usually fail
  - Cost savings immeasurable compared to cost of collecting the data
  - TW can't corroborate results
- Best practice
  - Ok to share 1st and 2nd generation data processing path
  - Plan for independent 3rd generation data analysis tools
  - If pressed to use a common set of tools, describe advantages of independent tools and return on investment of collecting the data
    - If unsuccessful, engage EN and TW management

FF11: Crosswind Takeoff and Landing Envelope Clearance

- Full crosswind clearance often delayed past IOC or the end of EMD
  - Many reasons:
    - Instrumentation, higher priorities, nature didn't cooperate, inappropriate planning
    - Easy enough to clear to 15 or 20 knots crosswind, MUCH harder to clear to 30
- Implement realistic plans
  - Do not commit to using EAFB main runway as only choice
    - Plan for alternates such as Mojave, Victorville, Palmdale
  - Do not depend on using EAFB lakebed runways
    - Especially for LO aircraft or configurations with low engine inlet(s)
  - Test plans need to recognize practical aspects of testing near design limits
  - When it's windy, the entire test team must be ready to GO
  - Assess test expertise

FF11: Crosswind Takeoff and Landing Envelope Clearance

- Avoid assumptions that lead to “artificial” impacts
  - Insure instrumentation design supports crosswind testing without reconfiguration
  - Advocate higher priority long before crosswind testing is the only remaining option
  - Do not require portable weather station when there is sufficient confidence in on-board GPS/INS to obtain wind values for data and KIO calls
  - Insure adequate pre-coordination with alternate test sites
    - Crash trucks, security, ground access, etc.

Concluding Thoughts

- Handbook intended as a living document
- Although Envelope Expansion centric, many LLs apply across disciplines
  - Need to expand to other disciplines
- Authors do not claim that methods described are the only way to go
  - Important to contemplate each topic in the context of a given project and develop creative solutions for that situation
- Not intended as contractor or program office “bashing”
  - By working to create a genuine team with foresight and willingness to act, it may be possible to avoid the “endless repetition of history”

Send Us Your Comments

Paul Sorokowski

paul.sorokowski@us.af.mil

Mike Garland

michael.p.garland@nasa.gov

APPENDIX B – MEASURING FLIGHT TEST PROGRESS ON LARGE SCALE DEVELOPMENT PROGRAMS

This page intentionally left blank.

MEASURING FLIGHT TEST PROGRESS ON LARGE SCALE DEVELOPMENT PROGRAMS

Charles H. Thornton, Capt, USAF*
Perry Lamy, Lt Col, USAF**

Abstract

Traditional methods of measuring flight test progress using flight hours and counting test points may not be adequate for large developmental flight test programs because of their size, complexity and political constraints. A tailored Cost/Schedule Performance Measurement System to measure the work scheduled and performed has been developed to improve the process of flight test planning, measurement, and control. Flight hours are used as the cost factor. A detailed computer automated program management network is used to integrate the technical requirements of test points between the various disciplines, the resources required to accomplish the test objectives, and the programmatic milestones which establish test priorities. Progress toward intermediate milestones, program milestones or certification of weapon system capabilities is credibly measured to provide a more meaningful report of flight test status to program office, service department, or DOD acquisition manager.

Nomenclature

AFWP	Actual Flight Hours Flown
BAC	Budget at Completion
BFWP	Budgeted Flight Hours for Work Performed
BFWS	Budgeted Flight Hours for Work Scheduled
E_F	Flight Hour Effectiveness Planning Factor
EAC	Estimate at Completion
ETC	Estimate to Complete
F_G	Flight Hours
F_N	Test Hours
FPI	Flight Performance Index
SPI	Schedule Performance Index
FV	Flight Hour Variance
SV	Schedule Variance
FTWO	Flight Test Work Order
MAC	Mean Aerodynamic Chord
ORD	Operational Requirements Document
PMB	Performance Measurement Baseline
TEMP	Test and Evaluation MasterPlan
TFPI	To Complete Flight Performance Index
SMM	System Maturity Matrix
STP	System Test Plan
WBS	Work Breakdown Structure

* Chief, Test Program Planning
Member

** Commander, 6520 Test Squadron
Member

Background

Why do projects fall behind?

- Poor Estimating techniques
- Confusing effort with progress (no mapping from effort to results)
- Pressure to deliver
- Progress is poorly monitored
- Tendency to add more resources to solve problem (time, funds, manpower)

Flight test programs traditionally track the progress of flight test using flight hours and test points. Flight hours alone are not representative of the completion of the test program. Test point completion may be adequate to satisfy overall C/SCSC criteria but it is not adequate for measuring progress and planing the effort because some test points are harder to get than others. For example, an envelope expansion test point may take 3-5 minutes each while a navigation test point could take 8-10 hours.

A test progress measurement system must adequately:

- Ensure that priority test objectives are accomplished to meet program objectives and milestones
- Establish insight into the flight hour requirement to meet the major test program milestones
- Measure the effective use of flight resources
- Identify critical test objectives based on resource availability, prerequisite testing, or external pressures such as software releases, congressional requirements, or operational certification

Approach

On large scale test programs, we must recognize all test points are not equal. Therefore a method of weighting test points is required. We have chosen test time on the resource as the weighting factor which provides a normalizing effect, i.e., a one hour test point has the same weight as 6 ten minute test points. This approach allows us to track the traditional flight hours and the test point hours allocated to various disciplines.

Test point hours are established by the engineering disciplines and reflect the data gathering time after the proper flight configuration has been established. It does not and should not include generic or common flight time for takeoff and landing, cruise time to ranges, in-flight configuration changes. Test point hours represent the direct cost in flight hours of performing the defined work and the generic or common flight time represent the indirect cost in flight hours.

This concept is similar to Cost/Schedule Control System Criteria (C/SCSC) used to monitor the performance of contracts. C/SCSC can be used with dollars or hours and all the elements are present to develop a credible performance measurement system.

To develop this capability, the following steps are required.

- Identify the Requirements.
- Define the Work
- Schedule the Work
- Baseline the Plan
- Monitor Performance and Analysis

Requirements

The requirements were identified to establish the work to be performed and the test priorities. We focused on maturing the system and user requirements. Significant test milestones were established in the Test and Evaluation Master Plan with the System Maturity Matrix. The maturity of the system must also support the operational demonstration and certification requirements derived from the Operational Requirements Document (ORD). Engineering design data requirements to verify the adequacy of design changes or software development also drive test priorities to support the maturity and operational requirements.

Defining the Work

Work Breakdown Structure

The work has been defined using a Work Breakdown Structure (WBS) into work elements by test category *(Figure 1)*. The categories represent Flight Test Work Orders (FTWO's) or groups of FTWO's which are further defined into individual test objectives. Work elements at the lowest level of the WBS, represent distinguishable test points and are linked to work packages developed in the PERT network.

Interrelationships of Work Packages

Once the work has been defined through the breakdown structure to the appropriate element, work packages are identified and interrelationships between other test objectives, hardware or software requirements are specified in a network *(Figure 3)*. This allows the network to schedule the work and establish priorities based on required test events. It also provides a projection of when program milestones will be complete and visibility into the latest period the aircraft must be configured

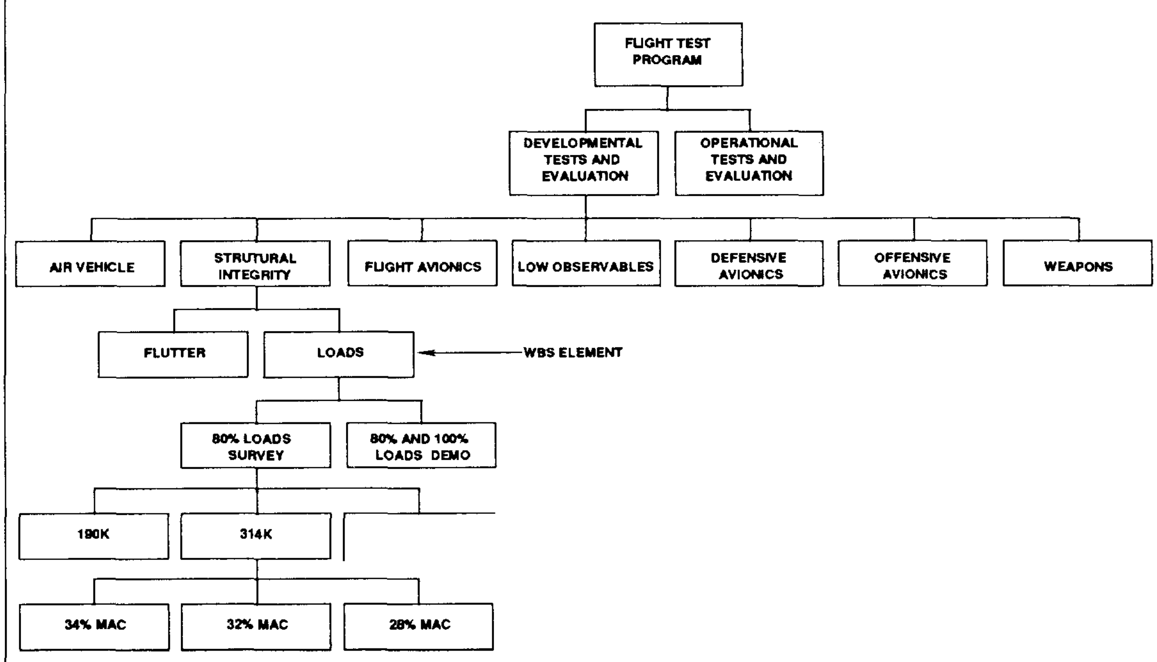


FIGURE 1. WORK BREAKDOWN STRUCTURE

to support required testing. The major priority driver will be the program milestones established by the requirement documents to meet System Maturity Milestones or deliver user capability.

Internal Work Package Definition

Each dedicated work package has been assigned a resource and a test hour allocation which represents a direct cost in hours (F_n) based on agreed to estimates for individual test points in the FTWO. Based on this

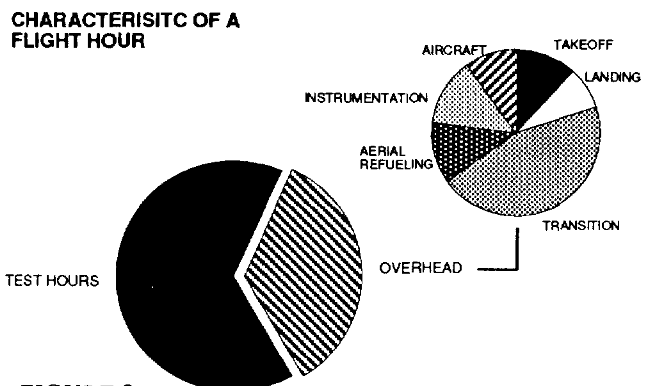


FIGURE 2.

allocation, an estimate can be made on the total test hour requirement (F_N) required to complete the program.

$$F_N = \Sigma F_n \quad (1)$$

The work identified in each work package represents the *direct* cost of obtaining data to satisfy the test objective but does not include the indirect cost of flight hours required to complete the work package. The *indirect* cost accounts for transition between test events, takeoff and landing, and aircraft reconfiguration in flight *(Figure 2)*. To account for this time, we have established an overhead planning factor. The direct and indirect cost for a work package makeup the flight hour requirement and represents the Budgeted Flight Hours of Work Scheduled (BFWS). The following relationship allows you to determine the BFWS for a work package from the test hour requirement:

$$BFWS = \frac{1}{E_f} * F_n \quad (2)$$

The total gross flight hour requirement for the test program can be represented as the *sum of the individual work package* requirements with the associated planning factor. It should be noted that the planning factor may vary for each work package based on the type of testing or air vehicle.

$$BAC = \Sigma BFWS \quad (3)$$

$$E_F = \frac{1}{BAC} * F_N \quad (4)$$

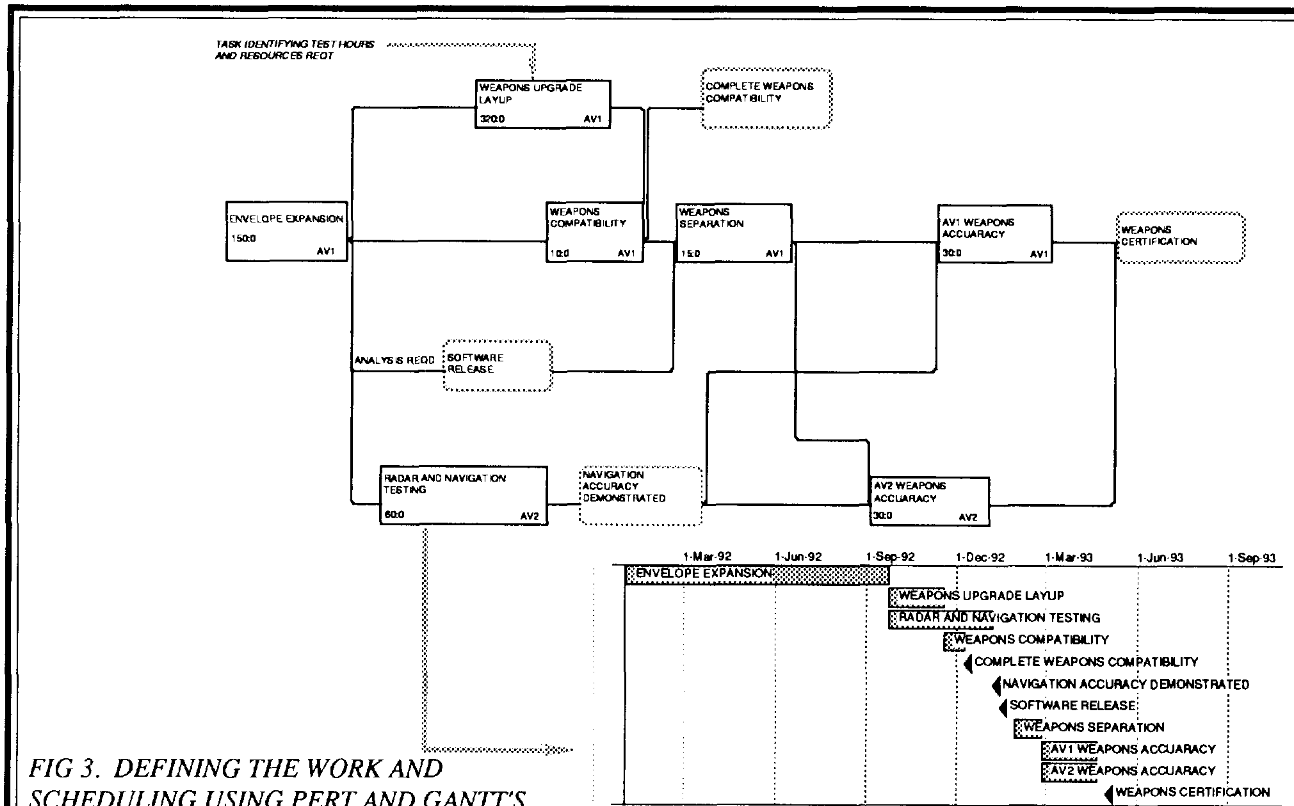


FIG 3. DEFINING THE WORK AND SCHEDULING USING PERT AND GANTT'S

Scheduling the Work

Now that we have an estimate of the gross (direct and indirect) flight hour requirements to complete the program, we must phase the work over time based on the expected flight rate of each air vehicle and when the various air vehicles will arrive from manufacturing. This is a top level approach which requires a commitment from management to provide the resources required to generate the air vehicle at a constant flight rate each month. Exceptions are made for planned "lay ups" or upgrades based on system maturity.

Baseline the Plan

The time phased plan *(Figure 4)* which is an accumulation of all the BFWS becomes the Performance Measurement Baseline (PMB). The sum of all BFWS plus management reserve is the Budget at Completion (BAC). At this point, it is very important that everyone involved in the mission planning and execution of the plan understand this baseline.

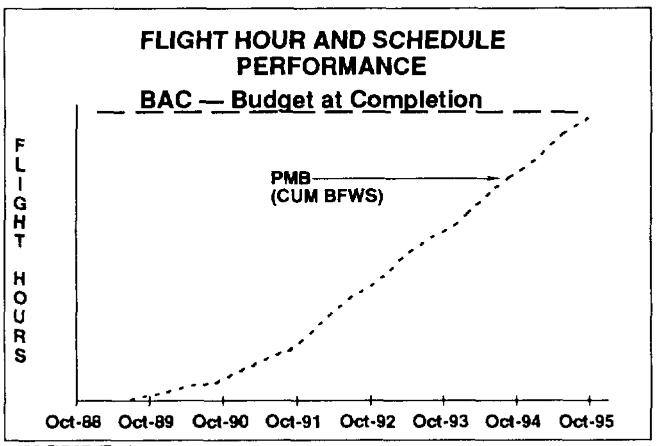


FIGURE 4

Track and Measure Progress

Now that we have established a Performance Measurement Baseline, we can track our progress against that baseline. We must track our ability to generate flight hours as planned and accomplish the work as planned when the aircraft is generated.

Actual Flight Hours

The actual flight time (AFWP) can be retrieved for each flight. The AFWP represents the actual cost in flight hours to complete work packages and gather the data to satisfy test objectives. When plotted with the PMB, we can determine if we are generating the flight hours required to accomplish the plan but we can not yet quantify the amount of scheduled work performed.

Earned Value

We have discussed developing a flight hour plan to accomplish the work and track actual flight hours against that plan. C/SCSC adds a third performance measurement called Earned Value. Earned Value is the physical value of the work actually accomplished relative to the actual flight hours flown and the test hours scheduled. A planning factor was used to develop the BFWS and at all times, the BFWP or Earned Value must be computed using the same planning rates as for BFWS.[1]

The individual test points and and associated test hours flown represent the physical work accomplished. The Earned Value (EV) can now be determined by applying the planning factor used to establish BFWS.

$$EV = \Sigma BFWP = \Sigma\left[\frac{E_n}{E_f}\right] \sim \frac{1}{E_F} * \Sigma F_n \qquad (5)$$

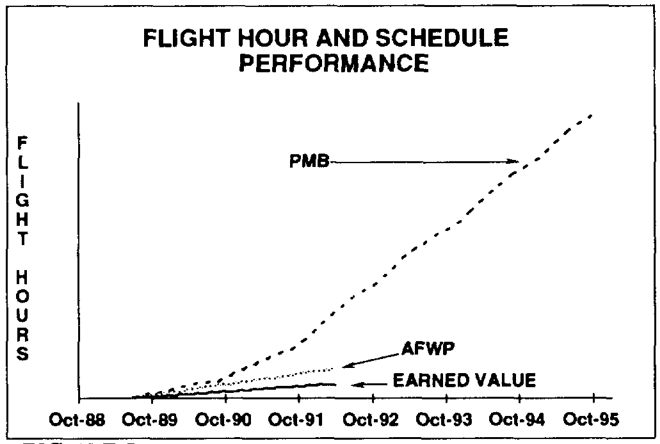


FIGURE 5

Performance Measurement and Analysis

Now we have a basis of comparing *actual vs planned* work, cost and schedule variances, efficiencies, and estimates at completion.

Suppose we have a flight test program with the following requirements.

- 19 month flight test program
- 1000 test hours
- 3 Flight test air vehicles

A detailed review of the work results in an estimated overhead planning factor of 35%. Management commits to a 65% effectiveness rate and will commit resources to fly 20 hours per month per air vehicle. A PMB is developed and baselined. The total flight hour requirement or BAC to complete 1000 test hours can be determined from Equation 4:

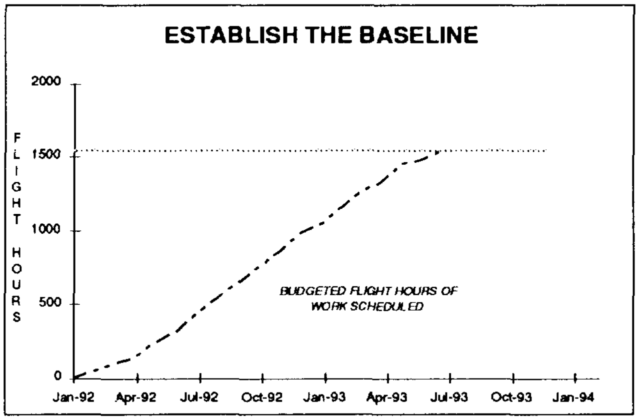


FIGURE 6

$$1540 \text{ Flight Hours} = \frac{1}{.65} * 1000 \text{ Test Hours}$$

Now lets move forward 11 months and look at the progress of the program. As of November 92, the following information has been tracked.

Test Hours Flown	392	ΣF_n
Flight Hours Flown	640	AFWP
Planned Flight Hours	862	BFWS

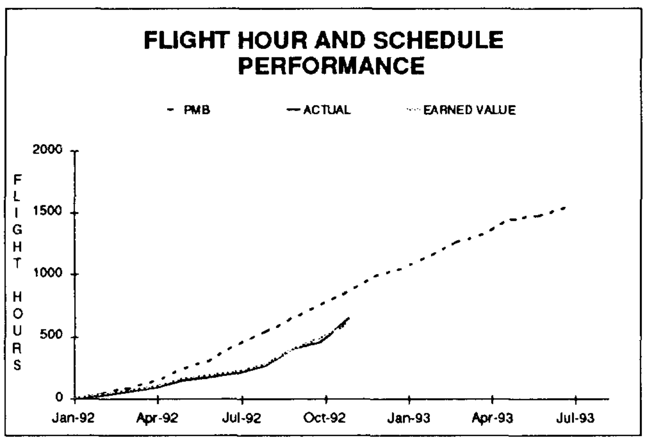


FIGURE 7

Immediately we can see the flight hours flown did not meet the amount planned. We can determine the average effectiveness of each flight hour by comparing the ratio of the test hours flown to the actual flight hours flown.

$$\text{Effectiveness} = \frac{\Sigma F_n}{AFWP} = \frac{392}{640} = 61.25\%$$

Using this relationship, we determine the actual effectivess of each flight hour has been 61.25% compared to the desired 65%. Using Equation 5, we can calculate the BFWP or earned value.

$$EV = \frac{1}{E_F} * \Sigma F_n = \frac{392}{.65} = 603 \text{ Flight Hours}$$

Variances from the Baseline

We can track our variance from the Performance Measurement Baseline or original flight hour estimate. The Flight Hour Variance gives us the difference between the number of *actual flight hours flown* and the *Budgeted Flight Hours* for the test hours accomplished. The Schedule Variance provides the deviation between the value of the work *accomplished* and the value of the work *planned to date*. It represents the variance from the Budgeted Flight Hours of Work Scheduled.

$$FV = BFWP - AFWP \quad (6)$$
$$FV = 603 - 640 = -37$$

$$SV = BFWP - BFWS \quad (7)$$
$$SV = 603 - 862 = -259$$

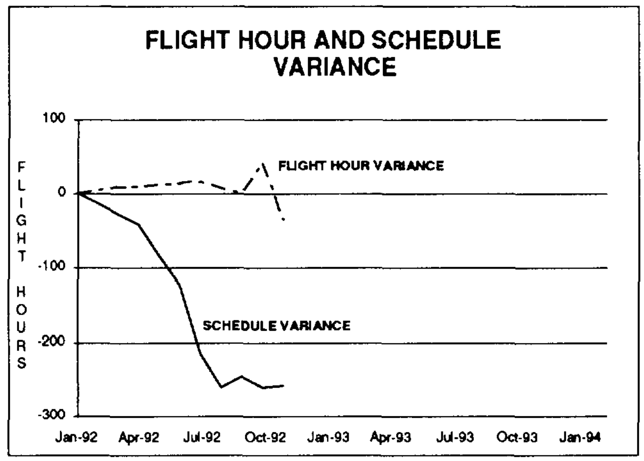


FIGURE 8

Variance thresholds can be established to trigger a comprehensive evaluation into the cause and breach criteria can be established to require a report by exception or rebaseline of the program.

Flight Performance Index (FPI)

We budgeted 603 flight hours to complete the 392 test hours which actually took 640 flight hours to accomplish. The ratio between the flight hours *budgeted* to accomplish the work and the *actual* hours expended represent a measure of efficiency. This measure is called the Flight Performance Index (FPI). An FPI of 1 means the use of each flight hour has been perfect, greater than 1 is exceptional, and less than 1 is not good. The FPI is calculated as follows:

$$FPI = \frac{BFWP}{AFWP} = \frac{603}{640} = .942 \quad (8)$$

A low FPI will require an evaluation of the mission and flight planning process, the health of the air vehicle, and instrumentation. These areas can cause an excessive

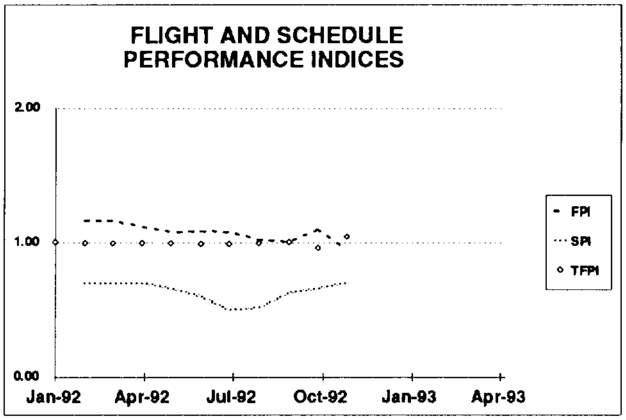


FIGURE 9

amont of non-productive flight time and cause a low FPI.

We should note the relationship between the FPI and the effectiveness rate.

$$\text{Effectiveness} = \text{FPI} * E_F = 61.25\% \quad (9)$$

<u>Schedule Performance Index (SPI)</u>

We can also assess our progress against our original plan, the PMB. When we compare the earned value to the BFWS we get a schedule efficiency. This tells us how well we are accomplishing our overall plan. Once again, 1 is perfect, greater than 1 is exceptional, and less than 1 is not good:

$$\text{SPI} = \frac{\text{BFWP}}{\text{BFWS}} = \frac{603}{862} = .70 \quad (10)$$

The SPI is inluenced by the mission execution, aircraft sortie generation, and the maturity of the air vehicle configuration to support flgiht test. This index ultimately will determine the lenght of the flight test program and the cost of maintining the manpower to support flight test. Reasons for not achieving the plan include generating insufficient flight hours, high air abort rates or a low FPI. The SPI would indicate a value of less than 1.0 if this is the case.

<u>Forecasting</u>

Determining the efficiency of accomplishing flight test goals and objectives allows us to also determine if the program will be completed within the allocated flight hours. Combined with the flight hour rate sustained, we can also estimate the impact on the overall schedule or flight test months.

Using our example, the *Estimate to Complete (ETC)* and the *Estimate at Completion (EAC)* may be determined by taking the difference between the original cost estimate and the Earned Value thus accounting for the flight performance index for efficiency:

$$\text{ETC} = \frac{\text{BAC - BFWP}}{\text{FPI}} = \frac{1540 - 603}{.942} = 995 \quad (11)$$

$$\text{EAC} = \text{ETC} + \text{AFWP} = 995 + 640 = 1635 \quad (12)$$

If our goal is to remain within the original BAC, we can also project what our FPI must be in order to achive that goal. The *To Complete Flight Peformance Index (TFPI)*

WHAT DOES IT MEASURE?

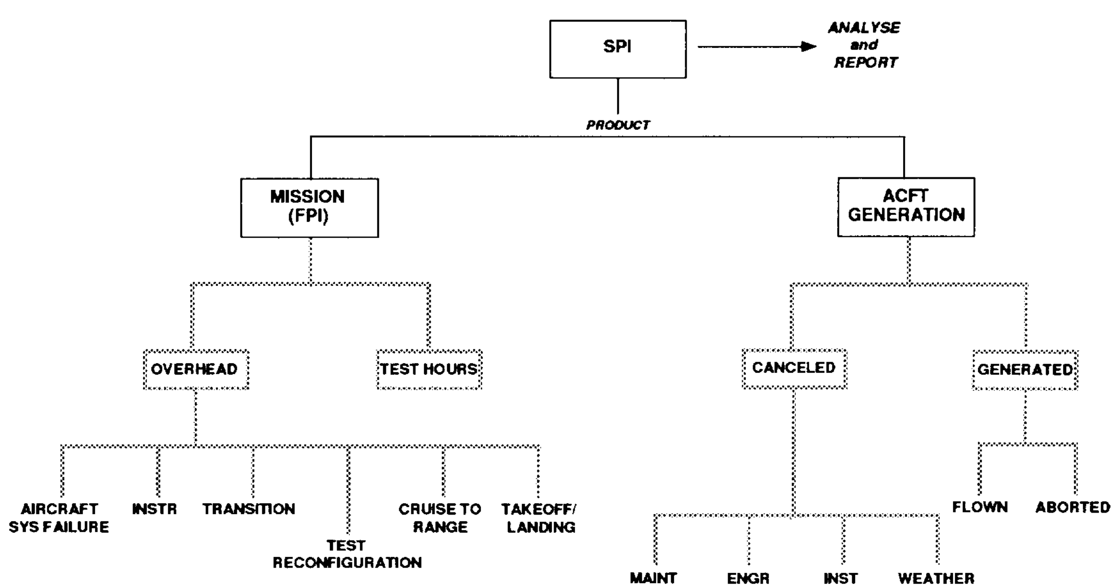


FIGURE 10

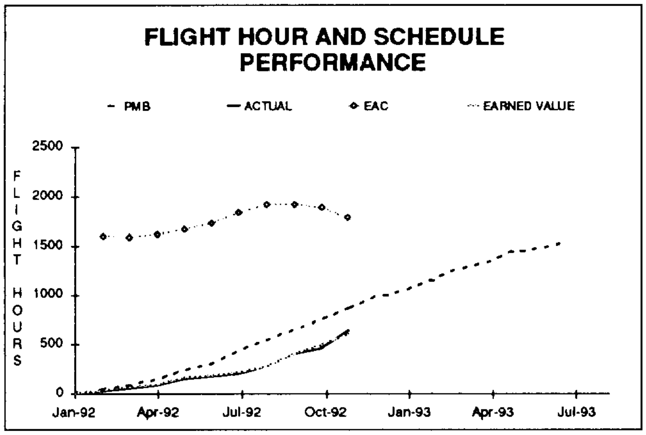


FIGURE 11

provides a reality check of the probabiliy of achieving the baseline once we have deviated from it:

$$TFPI = \frac{BAC - BFWP}{BAC - AFWP} = 1.04 \qquad (13)$$

In the above example, achieving the test objectives within the flight hour allocation seems reasonable, but the program is *259* flight hours behind schedule.

Conclusion

Flight test managers must use an accurate method of measuirng flight test progress and the efficiencies of meeting flight test objectives. Using this method, we can now quantify the causes of any inefficiencies under our control. The FPI and SPI are powerful indicators that quickly provide a managment reference on the status and efficiency of the flight hour program.

Using this measuring system will provide early indicators allowing you to accurately:

- Measure the result of action taken to increase the effective use of flight hours;
- Forecast additional flight hour requirements;
- Identify reductions in scope and
- Trade-off between higher flight rates and the additional resources required *now* against extending the program with associated cost.

This method combined with meeting specific program milestones will provide a more reliable assessment of flight test progress to the program office and DOD acquisition managers.

ENDNOTES

[1] DOD, *Cost/Schedule Control Systems Criteria Joint Implementation Guide*, 1 Oct 87, p. 3-15.

[2] Fleming, Quentin W., *Cost/Schedule Control Systems Criteria, The Management Guide to C/SCSC*, 1992

This page was intentionally left blank.

APPENDIX C – 411TH FLTS WIT AND DR CONCEPTS (EXAMPLE)

This page was intentionally left blank.

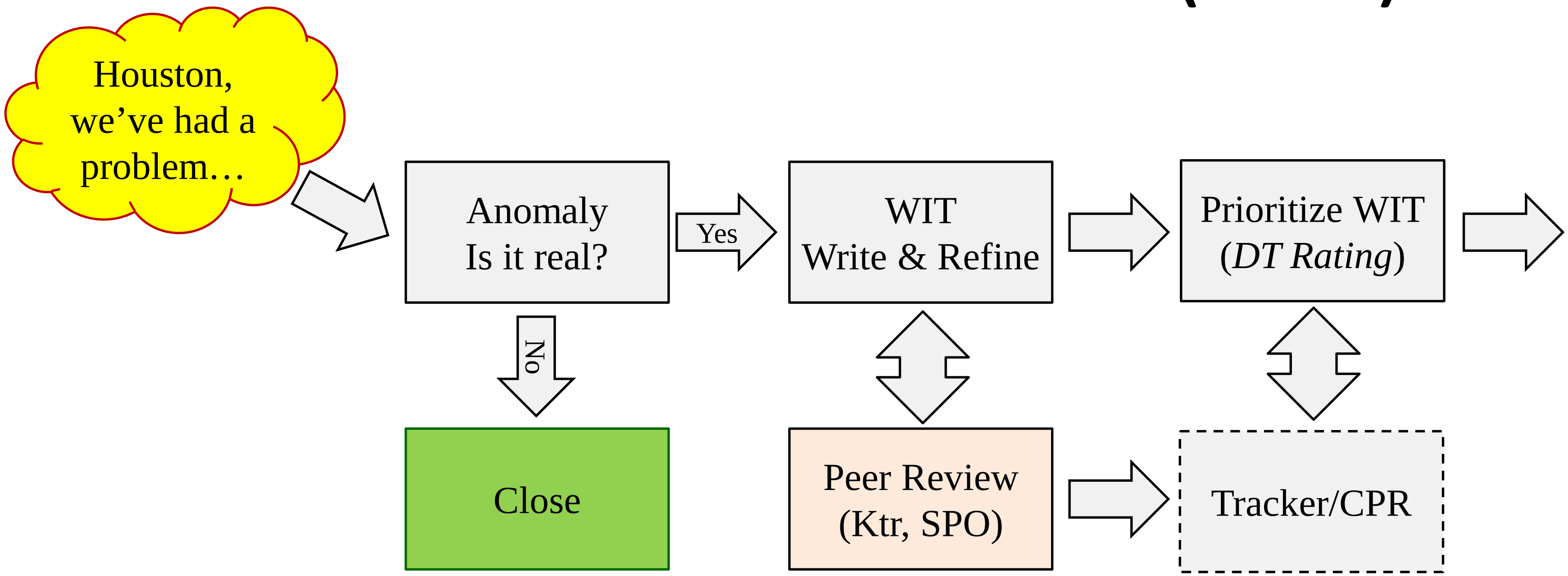
The Life of a 411 FLTS Watch Item (WIT)

- Sep 2016

411 FLTS WIT Background

- 411 FLTS had strong, positive relationships
  - SPO: Receptive to CTF perspective
  - AFOTEC/ACC/OT: Working together, sharing assets
  - Contractors: Integrated, receptive to CTF input
- Watch Item (WIT) DT Rating system contribution
  - Communication tool for gaining consensus
  - Focused Enterprise on what to fix to pass OT
  - Regular battle tempo to update the Enterprise
- Goal: Decision quality information as early as possible, no surprises for the Enterprise from DT

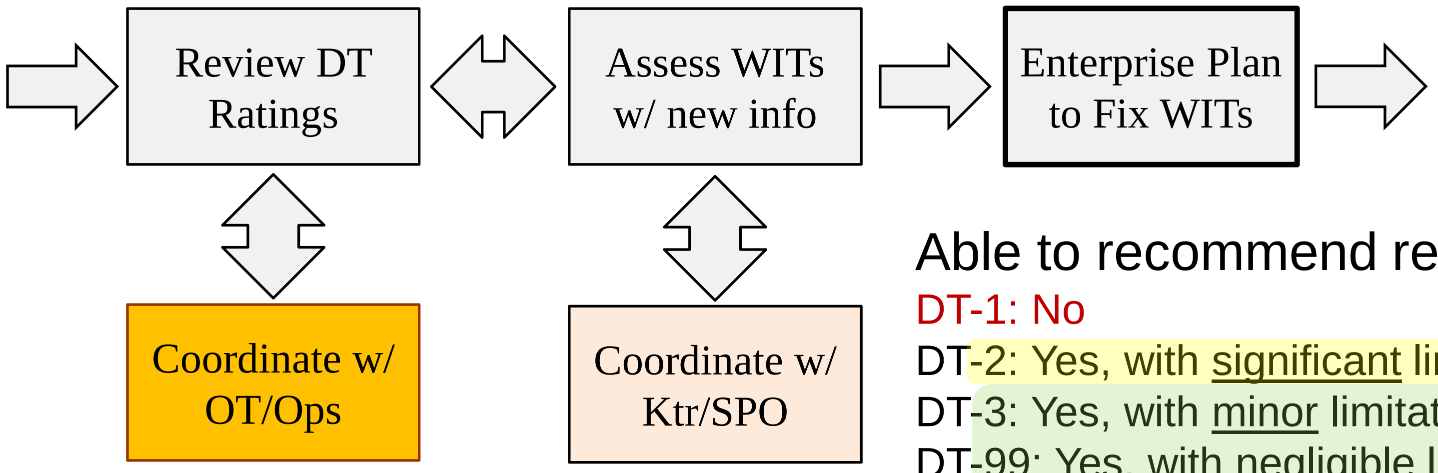
Birth of a Watch Item (WIT)



- Watch Items (WITs) get reviewed and clarified as soon as the CTF believes them to be a real problem
- They get linked to one or more contractor (Ktr) problem reports (Tracker/CPR) and refined as the team learns more about the problem
- They get rated by severity to help focus the team's effort

Note: A "Tracker" is just another type of CPR

Maturing of a WIT



Able to recommend release to OT?

DT-1: No

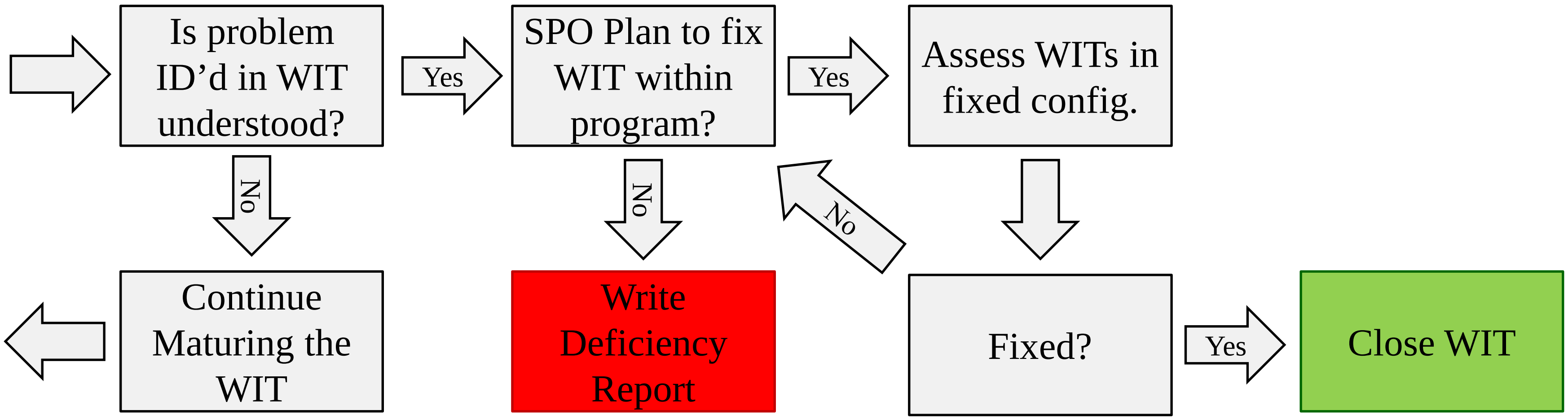
DT-2: Yes, with significant limitations

DT-3: Yes, with minor limitations

DT-99: Yes, with negligible limitations

- In an iterative process, the COMBINED Test Force reviews all the WITs to ensure appropriate rating
  - Coordinate with OT/Pilots and contractor/SPO technical experts
  - Revise rating based on the CTF assessment of the most complete information
- Potential enhancements still go through this process to determine if they are within scope of the contract

Death of a WIT



- By Air Force Instruction, a WIT can only be temporary
  - If fixed within the program, the WIT is closed – No need for a Deficiency Report
  - If there is no plan to fix a WIT or WITs remain at the end of a program, it must become a DR in JDRS

Notes:

- For a sustainment effort, "within the program" has meant "within a specific funded project to add capability"
- For a broader EMD-type effort, "within the program" could be at logical decision points, such as "prior to Milestone C decision" or "before IOC" or whatever makes sense for that project

EXAMPLES OF DT-RATING CHARTS

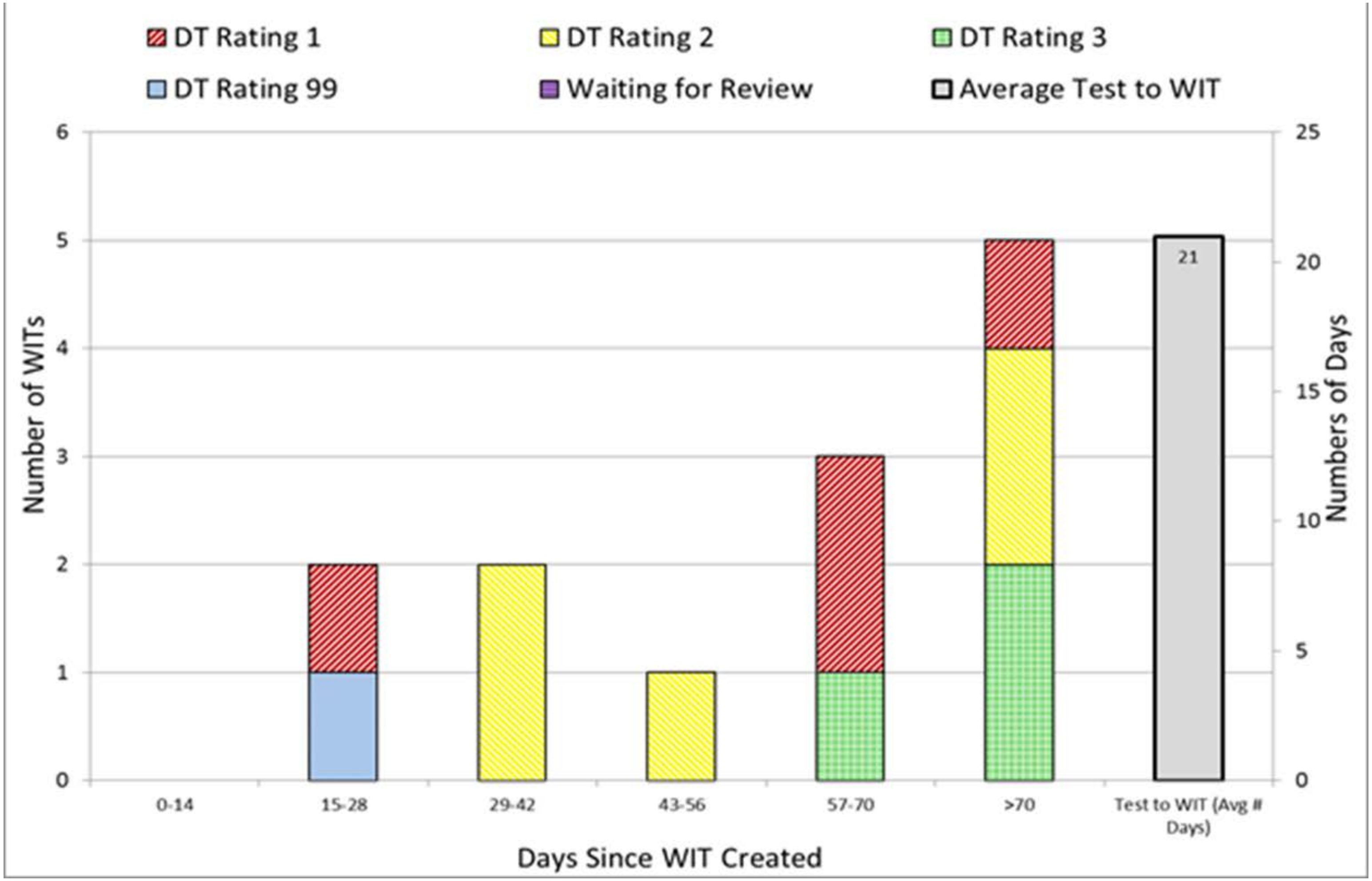
Program Assessment History & Prediction for Decision X

(as of date)	Configurations/dates			Total WITs/Rating			Total WITs without a planned fix		
	C1	C2	C3…	DT-1	DT-2	DT-3	DT-1	DT-2	DT-3
Capability 1	R	Y	G	10	20	30	0	2	3
Capability 2…	These would be hardware/software configurations as planned to drop before the decision time. For past configurations, the color indicates what the state was; for future configurations the color indicates when planned fixes drop into the program			This would be the number of DT-Ratings against the capability. When WITs are closed the number gets smaller.			This would be the number of DT-Ratings against the capability that do not have a confirmed plan to be fixed by the decision date		

For instance, the Capability 1 story is as follows:
The carrot indicates the configuration being tested. The C1 configuration is not recommended for release to OT because there is 1 out of 10 DT-1 rated WITs that the program has not fixed. The good news is that in the C2 configuration, there is either only DT-2 or less rated WITs or there is a program committed plan to have each DT-1 fixed for this capability, as a result the C2 configuration would be recommended for release to OT with significant limitations. Similarly, The C3 configuration is predicted to have either only DT-3 or less rated WITs or there is a program committed plan to have each DT-1 and -2 fixed for this capability, as a result the C2 configuration would be recommended for release to OT without significant limitations.

Backup slides use this color scheme to break each capability down to the WIT level and provide more status detail

Example Histogram Tracking of WITs Without Contractor Problem Reports



Status as of: 01/13/2016

Histogram Tracking of WITs Without Contractor Problem Reports

- WITs without a CPR for long time periods indicates inadequate attention
- Regardless of DT-rating, the histogram shows how long it is taking the CTF to produce a WIT and how long it is taking the KTR to start investigating a WIT.
- The intent is to ensure all problems get timely review and analysis so that priority WITs can get fixed first and faster.

Back-up Slides

WIT to Tracker/CPR Process

- Includes only problems ID'd during a CTF test, but may reference:
  - SIL/SIM problems with mission critical or safety of flight implications
  - Lab results may be tracked but do not create WITs
  - Lab results may inform tailored CTF test expectations
- CTF tests result in the following:
  1. Peer-to-peer review of unexpected test results
  2. Unexpected test results tracking
     - CTF creates a WIT database entry
     - Contractor creates a Tracker/CPR database entry
  3. CTF prioritizes WITs per DT Priority Definitions
     - Priorities include synergistic effects (e.g. the sum of certain DT-2 WITs may equate to a DT-1)
     - Bi-weekly (every other week) priority reviews & mid-level manager distribution
     - Alternate bi-weekly CTF Metric update
     - Periodic senior leader status briefing
  4. CTF watches deficiency until (within the project)
     - Fixed → CTF closes WIT or
     - Remains unfixed (or no plan to fix) → Deficiency Report process

Note: A "Tracker" is just another type of CPR

DT Priority Definitions

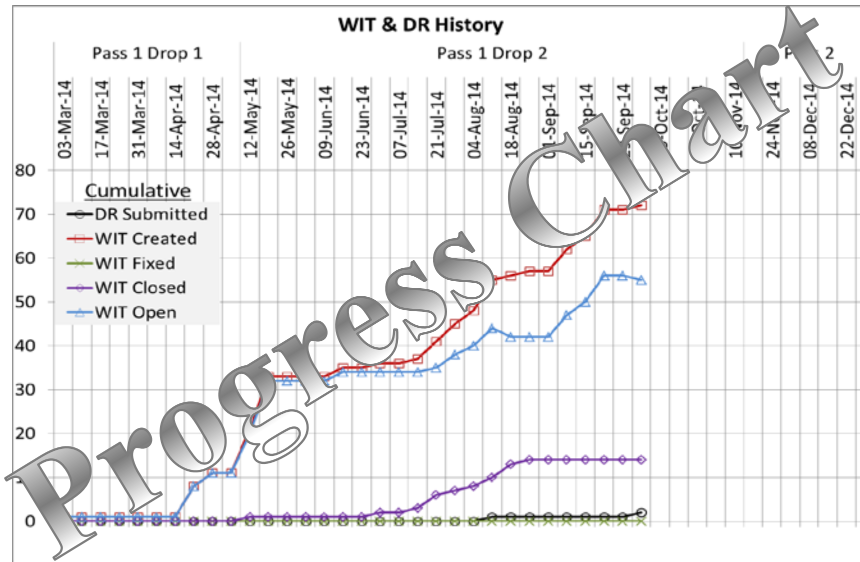
- DT Ratings – Anticipated CTF recommendation for OT or operations release rating
  - DT 1: Not able to recommend release
  - DT 2: Able to recommend release but with significant limitations or workarounds
  - DT 3: Able to recommend release with minor limitations or workarounds
  - DT 99: Able to recommend release with negligible limitations or workarounds
- DT Ratings include synergistic effects of WITs (e.g. some WITs when taken together make for a bigger problem)
- WITs include problems that effect only flight test so that programs understand how the problems alter test conduct
- WITs include enhancements (items that may not violate a hard requirement but could still be improved)

Project Status

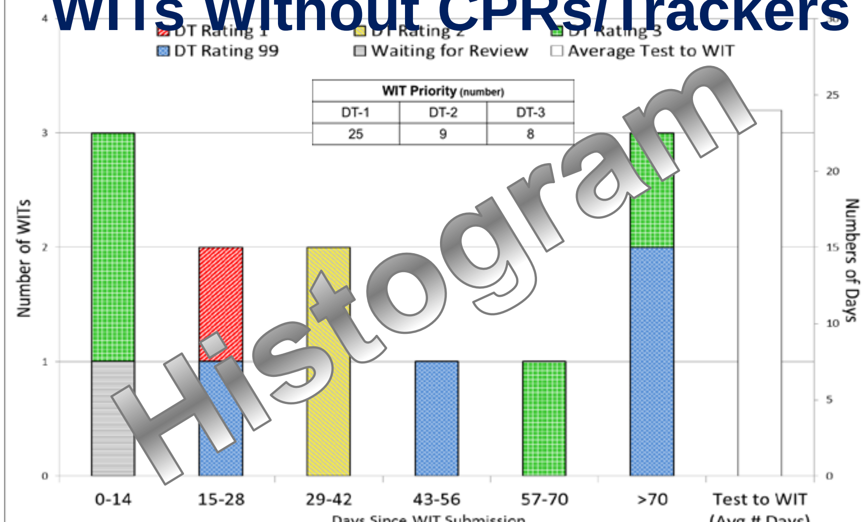
System Performance Assessment

Estimate* Now	Y
Estimate* At Completion	G

** These estimates may change if tests uncover problems, including the last test and associated evaluation.*



WITs Without CPRs/Trackers



Priority WITs

Table summarizing WITs Driving Assessment

List total number in each category & list titles of the biggest hitters.

Project Status

System Performance Assessment

(when evaluating ~~Incremental (I-Drop)~~ OFP)

Estimate* Now	**Y**
Estimate* At Completion	**G**

R — **Significant unexpected test limitations due to lack of or incorrect functionality in the OFP, or an unexpected un-fieldable degradation ~~to legacy capability~~. No fixes or workaround.**

Y — **Some unexpected test limitations due to lack of, or incorrect functionality in the OFP, or an unexpected degradation ~~of legacy capability~~ with some workarounds. Fix/capability planned in later ~~I-Drop~~ OFP or viable workaround exists.**

G — **No significant test limitations or unexpected degradation ~~to legacy capability~~.**

- **Estimate Now is based on performance to date of incremental capability and expected ~~legacy~~ capability.**
- **Estimate at Completion is based on performance assuming ID'd fixes are successful at delivery ~~of F1~~.**

***Caveat – This top-level assessment reflects a snapshot in time. In addition, the assessment may be based on partial or developing information. Users are advised not to extrapolate beyond the day the slide was created, as the assessment may change at any time.**

Project Status

System Performance Assessment

Estimate* Now	**Y**
Estimate* At Completion	**G**

R	**Not recommended for release. No planned fix or workaround.**
Y	**Recommend release with tolerable limitations or workarounds.**
G	**Recommend release, no significant limitations.**

WIT w CPR & CPR w/ Fix	WIT w CPR & CPR w/o Fix
Estimate Now If the program ends now, assumes CPRs are NOT implemented before fielding. WITs become DRs.	Estimate Now If the program ends now, WITs become DRs.
Estimate At Completion Assumes CPRs are implemented before fielding AND the CPRs correct the problems.	Estimate At Completion Assumes CPRs are NOT implemented before fielding.
WIT w/o a CPR (also no Fix)	
Estimate Now If the program ended now, WITs become DRs.	Estimate at Completion Same as Estimate Now.

"Fix" is shorthand for a Planned Fix within the program. This table explains how the Estimate definitions would be handled in each case.

Estimate Now is based on performance to date and assuming no future improvements. Estimate at Completion is based on performance assuming ID'd fixes are successful. At completion all open WITs will become DRs.

***Caveat – This top-level assessment reflects a snapshot in time. In addition, the assessment may be based on partial or developing information. Users are advised not to extrapolate beyond the day the slide was created, as the assessment may change at any time.**

WIT Priority

WIT Priority (number)					
DT-1	DT-2	DT-3	DT-99	NR	Total
17	7	3	3	1	31

- WIT Priority – Lists total open WITs by DT rating
  - DT Ratings – Anticipated CTF recommendation for OT or operations release rating

    DT 1: Not able to recommend release

    DT 2: Able to recommend release but with significant limitations or workarounds

    DT 3: Able to recommend release with minor limitations or workarounds

    DT 99: Able to recommend release with negligible limitations or workarounds

    NR: Needs Review
  - WIT DT Rating may change as research or additional data increases or decreases the resulting priority

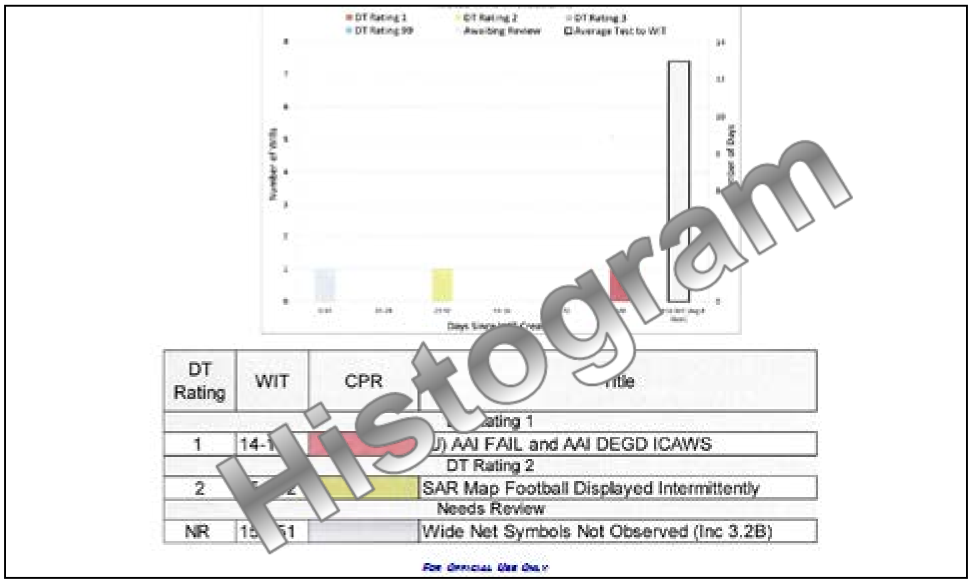
Priority WITs

Table summarizing WITs Driving Assessment

The Priority WIT List is lower left quadrant of the Project Status chart and part of the System Performance chart. The table lists the status of the highest priority WITs by the DT Rating priority (see DT Priority Definitions)

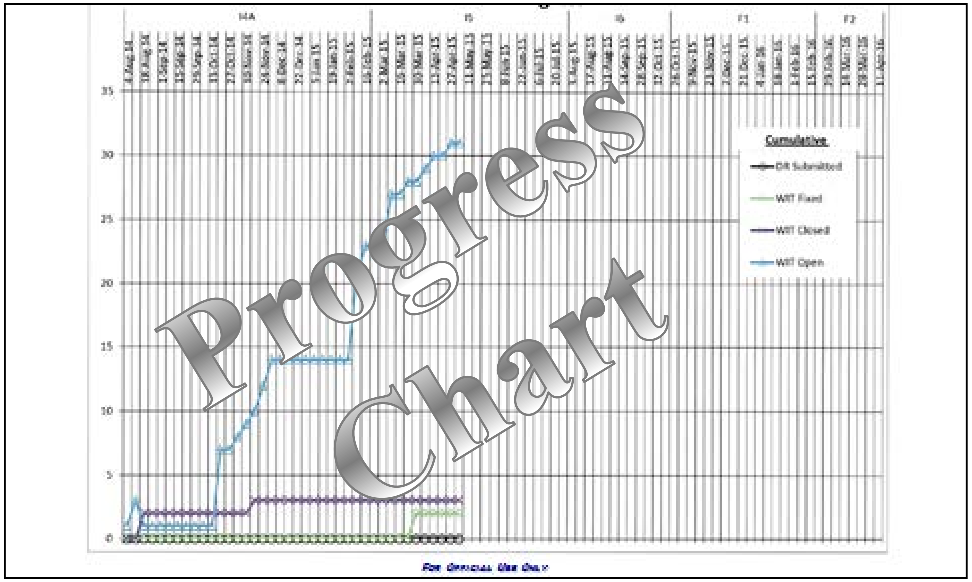
- **The following columns of the table lists a selection of the most significant WITs**
  - **DT Rating – see DT Priority Definitions slide**
  - **WIT – the control number of the watch item**
  - **CPR/Tracker – the control number(s) of the associated CPR or Tracker**
  - **Title – an unclassified title of the WIT**
  - **Fix Planned – the configuration planned to fix the problem (blank means a plan is not in place)**
  - **Fix Tested – Yes indicates that the planned fix has been tested.**
  - **Fix Verified – Yes indicates that the fix performed as planned, Analysis indicates that the data is available to be analyzed, and No indicates that the planned fix did not perform as expected.**
  - **Effectivity (if needed) – An X indicates to which aircraft configuration the WIT applies.**
- **Color of text – Red text signifies that a WIT is linked to the color status of the CTF's System Performance Assessment.**

WITs Without Trackers/CPR



- The WITs Without Tracker Chart is the Lower Left quad of the Project Status and also a separate WIT Without Tracker Chart.
- The histogram tracks the number of days that a WIT has been waiting for a Tracker.
  - WITs are removed from the histogram and table when a Tracker is assigned
  - When the program decides not to create a tracker/CPR the table lists “NONE”
- The second axis provides the average number of days from test to WIT submission.
  - The average is calculated from when the CTF observes a problem in ground or flight test to the date a WIT is entered into the database.
- The table provides a list of WIT titles that do not have Trackers or CPRs associated.

WIT Process Progress



- The Progress Chart is the Upper Right quad of the Project Status and the separate WIT Progress chart.
- Number of WITs or DRs is on the Y-axis.
- Date is on the X-axis.
- Legend definitions:
  - **DR Submitted** is the cumulative number of DRs submitted
  - **WIT Fixed** is the cumulative number of WITs that the CTF assesses to be fixed
  - **WIT Closed** is the cumulative number of WITs that do not require a DR.
  - This database status includes Closed Unsubstantiated (reject) or Closed Combined with another WIT that will be upgraded to a DR or was fixed.
  - **WIT Open** is the total number of WITs submitted minus DR Submitted, WIT Fixed, and WIT Closed.

This page was intentionally left blank.

APPENDIX D – OSD GUIDE FOR “INCORPORATING TEST AND EVALUATION INTO DEPARTMENT OF DEFENSE ACQUISITION CONTRACTS”

This page was intentionally left blank.

Incorporating Test and Evaluation into Department of Defense Acquisition Contracts

CLEARED
For Open Publication
OCT 24, 2011
Office of Security Review
Department of Defense
12-S-0150

October 2011

Office of the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation

Supersedes May 2009 Version

Published October 2011

Citation of this work should appear as follows: *Incorporating Test and Evaluation into Department of Defense Acquisition Contracts,* August 2011. Washington, DC: Office of the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation.

Office of Primary Responsibility:
Capabilities Development
Office of the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation
3040 Defense Pentagon, 5A1076
Washington, DC

The Capabilities Development office will develop and coordinate changes to the guide as required, based on policy changes and customer feedback. To provide feedback, please send comments by e-mail to DASD-DTE@osd.mil.

CONTENTS

FIGURES

TABLES

RECORD OF CHANGES

Change Number	Date of Change	Change

This page intentionally left blank.

1. INTRODUCTION

1.1. Purpose

This guide is designed to help the Department of Defense (DoD) and industry test and evaluation (T&E) professionals identify T&E items to consider for inclusion when drafting a statement of objectives (SOO), a statement of work (SOW), and a request for proposal (RFP), and during solicitation and contract execution. This guide presumes the reader has an understanding of T&E and the DoD systems acquisition processes as described in DoD Instruction 5000.02 (DoDI 5000.02) (Reference (a)) and the Defense Acquisition Guidebook (DAG) (Reference (b)), particularly, DAG Chapter 9, Integrated Test and Evaluation. This guide follows some of the content of the DoD Guide for Integrating Systems Engineering into DoD Acquisition Contracts (Reference (c)) and T&E topics and issues. This guide is for information purposes only.

This guide is structured to address generic T&E items common across DoD Components. Components may have specific T&E direction and guidance that each deems necessary for its acquisition programs. Most contracts begin at Milestone (MS) B – program initiation. However, a contract may be required prior to MS B for competitive prototyping. Programs may be required to implement acquisition strategies requiring a technology phase in which two or more competing teams will produce prototypes of the system or key system elements. Consequently, the SOO, SOW, and RFP are needed for prototyping contracts. These documents are essentially the same as those described in this guide for MS B.

The T&E guidance is based on programs that implement an acquisition strategy in which the development and testing have a single prime contractor. This is one of many DoD contracting strategies. Some project/system acquisitions will have different contracts. For example, Department of the Navy warship and combat system acquisition category programs may contract the engineering and production work to other Government and industry organizations for risk mitigation of the prime contract work. Regardless of the contract type, the important thing is to consider T&E requirements in the context of the contract. The program manager (PM) can tailor the T&E guidance to fit his or her particular situation or approach.

This guidance is based on the sequenced development process of a SOO, SOW, and the RFP leading to a contract. The underlying T&E considerations also apply to a rapid acquisition and fielding process, although the rapid process requires a much more focused T&E strategy and approach based on performance of key system capabilities and safety. The T&E strategy, including modeling and simulation (M&S), is an event-driven T&E approach linking key decisions in the system life cycle to knowledge from developmental and operational evaluations and outlines the test methodologies to obtain the data for evaluation. The T&E approach is an event-driven plan that identifies general or, when known, specific T&E techniques that contributes to capability maturation through discovery of performance levels and deficiencies including decision sequences and corrective action periods (CAPs). The strategy for T&E includes as much T&E information as is known at the time of development. The T&E strategy is captured in the approved test and evaluation strategy (TES) document/test and evaluation master plan

(TEMP) at milestones and focuses on the T&E events and activities expected in the technology development (TD) phase. The TEMP includes as much information as is known at the time of development. The TEMP is a Government document required prior to MS B and is typically not a contractual compliance document for inclusion in the RFP but is available in the program's document library for reference.

The primary theme to remember is that if a T&E item or requirement is not in the SOW, it probably will not be in the RFP, and if it is not in the RFP, it probably will not be in the contract. If it is not in the contract, *do not expect to get it!*

The T&E community consists of a broad range of personnel who perform a wide variety of T&E functions. When this guide refers to T&E personnel, ensure that the appropriate type(s) of T&E personnel with the appropriate T&E skills to provide the required support are cited. For example, when addressing the translation of critical technical parameters (CTPs) into contract specifications, this guide recommends that persons skilled in research, development, and T&E be assigned to write and/or review those parts of the contractual documents. When addressing contractor support needed for operational test and evaluation (OT&E), the OT&E personnel from the operational test agencies (OTAs) should be enlisted to write and/or review those parts of the contractual documents.

The "Lead for T&E" is a generic term referring to T&E personnel who lead the effort for T&E review, coordination, etc., for the integrated test team (ITT) or program office when T&E portions of contractual documents are being developed. The Lead for T&E may be one or several subject matter experts (SMEs) who bring specific T&E skills to the table.

1.2. Guide Organization

This guide contains the following four sections, organized to help the user focus on specific segments of the contract development process:

- **Section 1. Introduction**. This section covers the guide's purpose, organization, and definitions and includes an overview of the Defense Federal Acquisition Regulation Supplement (DFARS) (Reference (d)).
- **Section 2. Pre-Solicitation**. This section discusses the importance of including the T&E contracting requirements, including the T&E strategy and approach in the acquisition plan, TEMP, incentives, RFP/contract incentive structure, SOO, and ultimately in the SOW.
- **Section 3. Solicitation**. This section summarizes the source selection focus for those T&E items in the technical, management, cost, proposal risk, and past performance elements of the source selection. The section highlights proposal documents that evolve into the negotiated contract.
- **Section 4. Contract Execution**. This section addresses the transition to execution, contract oversight, and administration, and Defense Contract Management Agency

(DCMA) support. The section discusses the key actions immediately following contract award.

1.3. Definitions

Following are definitions for the principal terms used in this guide.

1.3.1. Contract

A contract is a mutually binding legal relationship obligating the seller to furnish the supplies or services (including construction) and the buyer to pay for them. It includes all types of commitments that obligate the Government to an expenditure of appropriated funds and that, except as otherwise authorized, are in writing. In addition to bilateral instruments, contracts include (but are not limited to) awards and notices of awards; job orders or task letters issued under basic ordering agreements; letter contracts; orders, such as purchase orders, under which the contract becomes effective by written acceptance or performance; and bilateral contract modifications. Contracts do not include grants and cooperative agreements (Federal Acquisition Regulation (FAR) 2.101 (Reference (e)).

1.3.2. Contract Data Requirements List (CDRL)

The CDRL (DD Form 1423) lists the contract data requirements authorized for a specific acquisition and becomes part of the contract. In addition, the CDRL may list packaging, packing, and marking requirements; delivery requirements; and work directed through special contract requirements. For more information on DD Form 1423 see http://www.dtic.mil/whs/directives/infomgt/forms/forsprogram.htm.

1.3.3. Contractor Role in T&E

The contractor will plan and execute the majority of design testing that transitions technology from science and technology efforts into functional capabilities desired by the military, as well as qualification testing of subcomponent parts and products from vendors that will make up the system delivered to the military. It will be necessary for Government testers to understand the contractor testing processes and methods to assess appropriate amount of visibility into those test activities as well as determine data collection and transfer that will benefit Government testers to avoid redundant or unnecessary testing. Experienced testers must determine cost/benefit ratios in visibility into proprietary activity and data transfer to the Government. Additionally, consideration must be given to near-end-state evaluations during operational testing (OT).

1.3.4. Data Item Description (DID)

A DID is a description of a data item that is to be put on the contract. Each data item will have its own DID. There are three types of DIDs: standard, tailored, and one-time. For more information, see http://www.dodssp.daps.dla.mil/assist.htm.

- **Standard DID**. A standard DID is one that is used “as-is.” A standard DID is used if it exactly describes the information requirement that needs to be put on contract.

- **Tailored DID**. A tailored DID is one in which not all of the requirements quoted in a standard DID need to be put on contract. The standard DID is "tailored down"; the scope of the DID is reduced by removing words, paragraphs, or sections. A DID can be tailored only by removing existing requirements from a standard DID. New requirements cannot be added to a standard DID. Many times, DIDs are tailored to accept a contractor's data format.
- **One-Time DID**. A one-time DID is used when a data requirement cannot be met by using a standard or tailored DID. These DIDs are written to acquire specific information on a specific contract.

1.3.5. Integrated Master Plan (IMP)

The IMP is an event-based plan consisting of a hierarchy of program events, with each event supported by specific accomplishments and each accomplishment associated with specific criteria to be satisfied for its completion.

1.3.6. Integrated Master Schedule (IMS)

The IMS is an integrated, networked schedule containing all the detailed discrete work packages and planning packages necessary to support events, accomplishments, and criteria of the IMP. A good source for more details on both the IMP and IMS is the Integrated Master Plan and Integrated Master Schedule Preparation and Use Guide (Reference (f)).

1.3.7. Lead for T&E

Lead for T&E is a generic term referring to appropriate T&E personnel who lead the development, writing, coordination, and review efforts for the ITT or program office for the T&E portions of contractual documents. The Lead for T&E may be one or several SMEs who bring specific T&E skills to the table.

1.3.8. Performance Work Statement (PWS)

The PWS is a SOW for performance-based acquisitions that describes the required results in clear, specific, and objective terms with measurable outcomes. See Subpart 2.101 of Reference (e).

1.3.9. Proprietary Right

Proprietary right is a broad term used to describe data exclusively owned by the contractor. These data could be intellectual property or financial data, for example. A contractor may use the term in a proposal to protect the contractor's sensitive information from disclosure, but the term is not a category of rights applicable to technical data, to include T&E data under all contracts.

1.3.10. Request for Proposal (RFP)

The RFP is a solicitation used in negotiated acquisition to communicate Government requirements to prospective contractors and to solicit proposals.

1.3.11. Statement of Objectives (SOO)

The SOO is the portion of a contract that establishes a broad description of the Government's required performance objectives. The SOO is a Government-prepared document incorporated into the solicitation that states the overall performance objectives. It is used in solicitations when the Government intends to provide the maximum flexibility to each offeror to propose an innovative approach. See Subpart 2.101 of Reference (e).

1.3.12. Statement of Work (SOW)

The SOW is that portion of a contract that establishes and defines the work to be performed by the contractor, and it may incorporate specifications, DIDs, or other cited documents. The SOW should be consistent with all "promises or claims" made in the proposal. A very good reference for SOOs, SOWs, and PWSs is the Defense Acquisition University (DAU) online Continuous Learning Module (CLM) 031, "Improved Statement of Work" (Reference (g)).

1.3.13. System Performance Specification (SPS)

The SPS or equivalent contents will be incorporated into the contract. The SPS describes the operational characteristics desired for an item without dictating how the item should be designed or built. The Joint Capabilities Integration and Development System (JCIDS) documents (i.e., capability development document (CDD), operating and enabling concepts) are the basis in developing the system specification. These documents are key to developing sound contractual documents. A complete understanding of the system, verifying system performance, and validating T&E results will ultimately be based on meeting JCIDS requirements.

1.3.14. Test and Evaluation Master Plan (TEMP)

The TEMP documents the overall structure and objectives of the T&E program. It provides a framework to generate detailed T&E plans and documents schedules and resource implications associated with the T&E program. The TEMP identifies the necessary developmental test and evaluation (DT&E), OT&E, and live-fire test and evaluation (LFT&E) activities. It relates program schedule, test management strategy and structure, and required resources to critical operational issues (COIs), CTPs, objectives, and thresholds documented in the CDD, evaluation criteria, and milestone decision points. The TEMP does not relieve the contractor of any contractual obligations. It serves as an indicator of Government expectations and should complement, not contradict, specifications and contractual language. The Government TEMP should be shared with industry as appropriate. Sharing the TEMP pays dividends and should be a common practice as appropriate to contractual T&E responsibilities (e.g., a single prime contractor responsible for all T&E).

1.3.15. Test and Evaluation Strategy (TES)

The TES is an early T&E planning document that describes the T&E activities starting with TD and continuing through engineering and manufacturing development (EMD) into production and deployment. Over time, the scope of this document will expand. The TES will evolve into the TEMP due at MS B. The TES describes, in as much detail as possible, the risk-reduction efforts across the range of activities (e.g., M&S, DT&E, OT&E, etc.) that will ultimately produce a valid evaluation of operational effectiveness, suitability, and survivability before full-rate production and deployment. It is a living document and should be updated as determined by the T&E working integrated product team (WIPT) during the TD phase. Its development will require early involvement of testers, evaluators, and others as a program conducts pre-system acquisition activities, especially prototype testing. The TES should be consistent with and complementary to the systems engineering plan (SEP).

1.3.16. Title 10, United States Code (U.S.C.)

Title 10 (Reference (h)), section 2399, Operational Test and Evaluation of Defense Acquisition Programs, paragraph (d), Impartiality of Contractor Testing Personnel, states that in the case of a major defense acquisition program, as defined in Reference (h), sections 139, 2399, 2430, and 2302(5), no person employed by the contractor for the system being tested may be involved in the conduct of OT&E, establishing OT&E criteria, or OT&E evaluation. The contractor can be tasked to provide technical understanding of test incidents, logistics support and training, support to test failure analysis, and unique software and instrumentation support. The limitation does not apply to the extent that the Secretary of Defense plans for persons employed by that contractor to be involved in the operation, maintenance, and support of the system being tested when the system is deployed in combat.

NOTE: System contractors are those who design and build the system, and support contractors are those who work for the Government in support of the acquisition and T&E of those systems.

System contractors may be beneficial in providing logistic support, test failure analyses, and software and instrumentation support that could increase the value of unprocessed OT&E data. Clear explanations of how system contractor support will be used and the mitigation of possible adverse effects must be described in the TEMP and OT&E plans to ensure no violation of the prohibitions in title 10, section 2399. Consider using the system contractor capabilities and skills in the following specific areas to support dedicated OT&E:

- Performing maintenance and support actions of the same type that the system contractor would be expected to perform as part of interim contractor support of contractor logistics support when the system is fielded.
- Conducting and reporting analyses of test failures to assist in isolating causes of failure but excluding participation in data scoring and assessment conferences.
- Providing and operating system-unique test equipment, test beds, and test facilities that may include software, software support packages, instrumentation, and instrumentation support. Full aircraft mission simulator systems are examples.
- Providing logistics support and operator training as required in the event such services have not yet been developed and are not available from the Military

Department or Defense Agency having responsibility for conducting or supporting OT&E.

- Providing data generated prior to the conduct of OT&E, if deemed appropriate and validated by the OT organization, to ensure that critical issues are sufficiently and adequately addressed.

1.3.17. Work Breakdown Structure (WBS)

The WBS is a fundamental project management technique for defining and organizing the total scope of a project, and delineates and segregates the technical elements to report costs to support technical management decisions and progress. A well-designed WBS describes planned outcomes instead of planned actions. The WBS needs to be consistent with the T&E program and the way in which it is conducted, or it may be difficult to evaluate. A very good reference for WBS information is the DAU online CLM 013, "Work Breakdown Structure (WBS)" (Reference (g)).

1.4. Defense Federal Acquisition Regulation Supplement (DFARS)

1.4.1. Using DFARS

Guide users are not expected to have the same knowledge as contracting officers (KOs) but should understand the purpose of DFARS and where to look for specific guidance and information. DFARS and a Service's or Defense Agency's contracting supplement provide specific clauses that must be included in the contract, and they may identify items for delivery. What is expected to be delivered is the main T&E focus, especially contractual language on proprietary/intellectual rights and data access and sharing.

1.4.2. DFARS Requirements

The DFARS remains the source for regulation and implementation of laws as well as DoD-wide contracting policies, authorities, and delegations. In other words, DFARS will answer these questions: *What is the policy?* and *What are the rules?* The DFARS Procedures, Guidance, and Information Website (Reference (d)) connects the acquisition community to available background, procedures, and guidance and answers these questions: *How can I execute the policy? and Why does this policy exist?*

1.4.3. Federal Acquisition Regulation (FAR) Part 16

FAR Part 16 (Reference (e)) , Service Supplements, and Individual Service Award Fee Guides provide additional information on types of contracts and incentives that may be used (FAR Subpart 16.4 (Reference (e)); DFARS Subpart 216.4 (Reference (d)); Army Federal Acquisition Regulation Supplement Subpart 5116.4; Air Force Federal Acquisition Regulation Supplement Subpart 5316.4; Air Force Award Fee Guide (Reference (i)); Army Award Fee Guide (Reference (j)); Navy/Marine Corps Award Fee Guide (Reference (k)). To search specific FARs go to http://farsite.hill.af.mil.

1.5. Acquisition Process

This guide focuses on contract development leading to contract award. Traditionally, program designation and contract award are at MS B. However, regardless of the acquisition phase, some contracts may be awarded prior to MS B, and the T&E contractual considerations described in this guide still apply. The five major phases of the Government acquisition process are defined in DoD Directive (DoDD) 5000.01 (Reference (l)), DoDI 5000.02 (Reference (a)) and the DAG (Reference (b). Figure 1-1 depicts the current Defense Acquisition Management System.

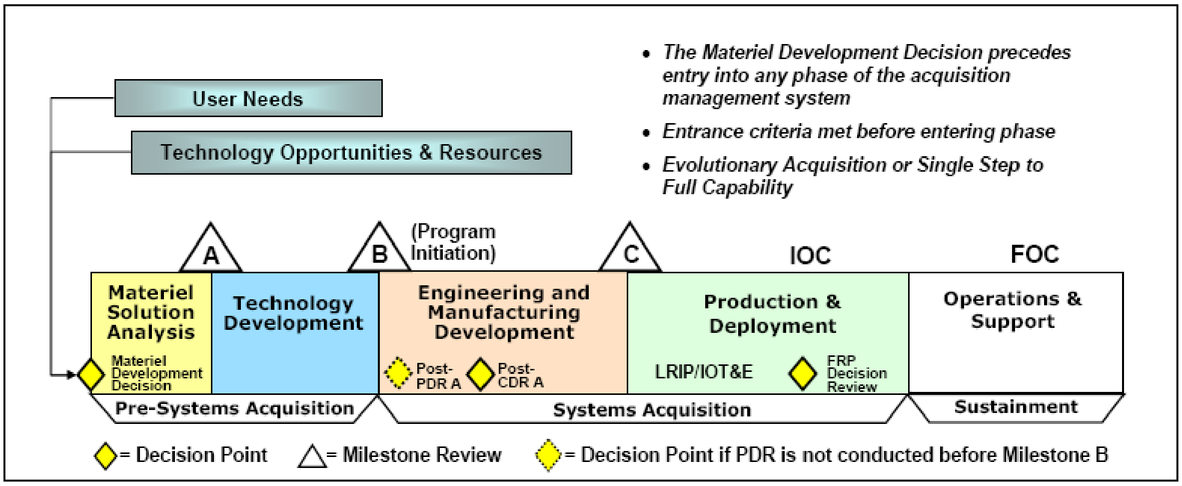


Figure 1-1 The Defense Acquisition Management System

Figure 1-2 is a simplified illustration of the above acquisition process depicting the associated contracting steps. It begins when the Warfighter identifies the need (See Chairman of the Joint Chiefs of Staff Instruction 3170.01G (Reference (m))) to the acquisition activity, which then translates that need into a requirement and purchase request. The KO solicits offers from industry and awards a contract. In the final step, the contractor closes the loop by delivering supplies and services that satisfy the Government need. Be aware that there may be a separate RFP for each phase of the acquisition process (e.g., RFP for TD, RFP for EMD, RFP for low-rate initial production, and RFP for production). During acquisition planning, primary responsibility rests with the acquisition activity.

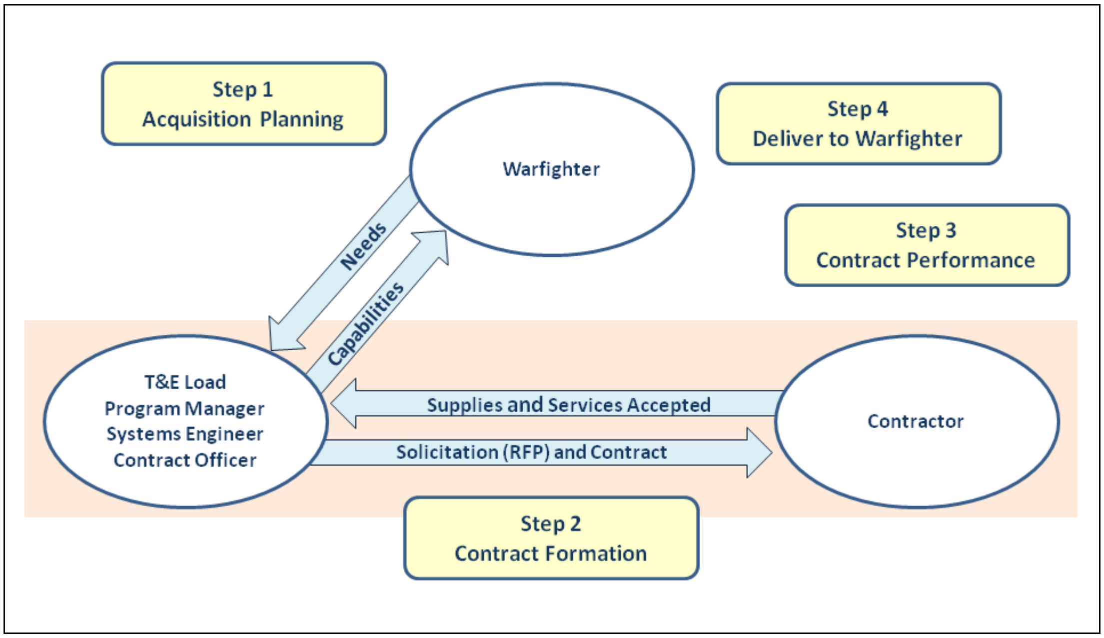


Figure 1-2 Simplified Government Acquisition Process

Acquisition planning is the process of identifying and describing contract requirements and determining the best method for meeting those requirements (e.g., business, program acquisition strategy (AS)), including solicitations and contracting. Acquisition planning focuses on the business and technical management approaches designed to achieve the program's objectives within specified resource constraints. The AS usually drafted in the TD phase of acquisition, is required and approved by the Milestone Decision Authority and provides the integrated strategy for all aspects of the acquisition program throughout the program life cycle. Earlier developmental activities are guided by the technology development strategy.

The TES and then the TEMP provide the approach on the content, management, and focus of the T&E aspects of the acquisition program. The acquisition plan provides more specific plans for conducting the acquisition and is approved in accordance with agency procedures, see Part 7 of Reference (e). A source selection plan (SSP) specifies the source selection organization, evaluation criteria, and procedures, and is approved by the KO or other Source Selection Authority. All of these documents guide RFP development. Other companion program artifacts include the capabilities documents (initial capabilities document (ICD), CDD, and the capability production document (CPD)); risk management plan (RMP); technology readiness assessment; information support plan; SEP; product support strategy; test plan; and support and maintenance requirements. A good source for policy and guidance is DAU's Acquisition Community Connection Practice Center Website (Reference (g)).

The program team must have strong technical, contracting, and T&E leadership as the program moves through its steps in contract formulation and execution. It is imperative to have the KO involved in the program acquisition planning process as early as possible.

1.6. Contracting Process

The PM, chief or lead systems engineer (SE), KO, Lead for T&E, and combat developer must work together to translate the program's requirements document, AS document or acquisition plan, and associated technical documents into a cohesive, executable contract, as appropriate. Table 1-1 identifies some typical acquisition process activities, starting from requirements identification through contract close-out, and capturing lessons learned and the role of the Lead for T&E who provides the T&E input, review, and coordination.

A wide selection of contract types is available to the Government and contractors to provide needed flexibility in acquiring the large variety and volume of supplies and services required by Defense Agencies. See Part 16 of Reference (e) for further information. Contract types vary according to:

- The degree and timing of the responsibility assumed by the contractor for the costs of performance.
- The amount and nature of the profit incentive offered to the contractor for achieving or exceeding specified standards or goals.

The contract types are grouped into two broad categories: fixed-price contracts (see Subpart 16.2 of Reference (e)) and cost-reimbursement contracts (see Subpart 16.4 of Reference (e)). The specific contract types range from firm-fixed-price, in which the contractor has full responsibility for the performance costs and resulting profit (or loss), to cost-plus-fixed-fee, in which the contractor has minimal responsibility for the performance costs and the negotiated fee (profit) is fixed. In between are the various incentive contracts (see Subpart 16.4 of Reference (e)), in which the contractor's responsibility for the performance costs and the profit or fee incentives offered are tailored to the uncertainties involved in contract performance.

Table 1-1 Acquisition Process Activities and the T&E Role

Typical Acquisition Process Activities	T&E Role
1. Identify overall procurement requirements and associated budget.	Lead for T&E determines testability of requirements and describes the Government's T&E needs and any constraints on the procurement from the program-related requirements provided by the PM.

Table 1-1 Acquisition Process Activities and the T&E Role

Typical Acquisition Process Activities	T&E Role
2. Identify T&E actions required to successfully complete T&E and performance milestones.	Lead for T&E defines the T&E strategy and approach and required T&E efforts. In consultation with or at the direction of appropriate T&E personnel, describes the Government's T&E needs consistent with the program's AS or acquisition plan. This effort should include defining contractor and Government testing, identification of test and training ranges of the test equipment and facilities, capabilities designated by industry and academia, unique instrumentation, threat simulators, targets, and M&S. Certain test events such as initial operational test and evaluation (IOT&E), interoperability certification, information assurance (IA), DoD Information Assurance and Certification Accreditation Process (DIACAP) certification and accreditation (C&A), and independent verification and validation may have to involve independent SMEs.
3. Collaborate on acquisition and T&E strategies.	The PM, combat developer, and appropriate T&E personnel collaboratively develop the acquisition and T&E strategies so that users' capability-based operational requirements (i.e., CDD, concept of operations (CONOPS)) are correctly translated into accurate contractual terms and actions that give the highest probability of successful outcome for the Government. Contracted events must provide for sufficient time to execute all regulatory and statutory T&E activities and reporting.

Table 1-1 Acquisition Process Activities and the T&E Role

Typical Acquisition Process Activities	T&E Role
4. Identify the reliability, availability, and maintainability (RAM) requirements and the need for a reliability program plan (RPP).	PM, SE, and Lead for T&E identify the RAM and RPP requirements for a robust RAM program, which includes reliability growth planning as an integral part of product/system design, development, and T&E consistent with technical maturity and the SEP. In addition, in accordance with Directive-Type Memorandum (DTM) 11-03 (Reference (n)), the sustainment characteristics of the materiel solution resulting from the analysis of alternatives (AoA) and the CDD, sustainment key performance parameter (KPP) thresholds will be translated into reliability and maintainability (R&M) design requirements and contract specifications. The strategies shall also include the tasks and processes to be stated in the RFP that the contractor will be required to employ to demonstrate the achievement of reliability design requirements. The tasks and processes will be compared against the plan (track/plan). Consider elements of T&E necessary for decisions points that will best balance RAM maturity with capability thresholds and objectives.
5. Perform market research to identify potential sources.	PM and Lead for T&E identify programmatic and T&E information needed and assist in evaluating the search results for each area. See Part 10 of Reference (e) for sources of market research, including trade studies, limited demonstration test results, and procedures. Small businesses must be considered.
6. Identify Human systems integration (HSI) and usability test criteria.	PM and Lead for T&E, in coordination with HSI SMEs, develop test criteria for HSI requirements explicitly stated in capability requirements, or derived from capability requirements and HSI guidance in MIL-STD-46855A (Reference (o)), MIL-STD-1472F (Reference (p)), and related DoD and DoD Component guidance. HSI may impact, and be impacted by, requirements and specifications in the areas of manpower and personnel planning; training; environmental, safety, and occupational health provisions; human factors engineering; and survivability and habitability provisions.

Table 1-1 Acquisition Process Activities and the T&E Role

Typical Acquisition Process Activities	T&E Role
7. Document the role of M&S.	PM, with the Lead for T&E, identifies the role M&S will contribute to the acquisition process. This effort should be consistent with the engineering plan for M&S. Address the need for an M&S support plan if required per DoD component direction.
8. Prepare a purchase request.	PM and Lead for T&E ensure that the specific programmatic and T&E needs are defined clearly. Consider the needs for testing commercial off-the-shelf (COTS) systems as well as any possible contractual implications regarding testing associated with Part 12 of Reference (e). A purchase request should include product descriptions; priorities, allocations, and allotments; architecture; COTS, Government-furnished information (GFI), or Government property or equipment; IA and security considerations; and required delivery schedules.
9. Identify acquisition streamlining approach and requirements.	The program team works together to ensure that FAR and DFARS requirements are met while tailoring the acquisition strategy and approach. The PM is owner of the program acquisition strategy and planning. The Lead for T&E develops and reviews (and PM approves) the T&E strategy and approach with the PM and lead engineer. Acquisition streamlining approach and requirements include budgeting and funding, contractor versus Government performance, management information requirements, environmental and safety considerations, offeror expected skill sets, and milestones. These are addressed in the AS document or acquisition plan.

Table 1-1 Acquisition Process Activities and the T&E Role

Typical Acquisition Process Activities	T&E Role
10. Determine contractor OT&E support.	In conjunction with the OTAs, the PM and Lead for T&E will define the degree of contractor support to be provided for OT&E. There are five permissible types of contractor OT&E support: (1) performing maintenance and support actions of the same type that the system contractor would be expected to perform as part of interim contractor support or contractor logistics support when the system is deployed in combat; (2) conducting and reporting analyses of test failures to assist in isolating causes of failure (but excluding participation in data scoring and assessment conferences; (3) providing and operating system-unique test equipment, test beds, and test facilities that may include software, software support packages, instrumentation, and instrumentation support; (4) providing logistics support and training as required in the event that such services have not yet been developed and are not available from the Military Department or Defense Agency responsible for conducting or supporting the OT&E; and (5) providing data generated prior to the OT, if deemed appropriate and validated by the independent OTA, to ensure that critical issues are sufficiently addressed.
11. Plan the requirements for the contract SOO/SOW specification, and T&E reviews in support of the technical reviews, test readiness reviews (TRRs), certifications for OT&E readiness, DIACAP C&A, acceptance requirements, and schedule.	Lead for T&E is responsible for developing the T&E contents of the SOO/SOW and supporting the technical reviews, TRRs, certifications for OT&E readiness, and DIACAP C&A.
12. Plan and conduct Industry Days as appropriate (See section 2.5).	PM and Lead for T&E support the KO in planning the meeting agenda to ensure that T&E needs are discussed.

Table 1-1 Acquisition Process Activities and the T&E Role

Typical Acquisition Process Activities	T&E Role
13. Establish contract cost, schedule, and performance reporting requirements. Determine an incentive strategy and appropriate mechanism (e.g., incentive/ award fee plan and criteria).	Lead for T&E provides resource, schedule, and performance estimates by developing the T&E portion of the WBS or work package based on preliminary system specifications; determines T&E event-driven criteria for key technical and readiness reviews; and determines what T&E artifacts are baselined. The PM, Lead for T&E, and lead engineer advise the KO in developing the metrics/criteria for an incentive mechanism.
14. Identify T&E data requirements.	Lead for T&E identifies all T&E CDRL intellectual property requirements, if any, and T&E performance expectations. This includes defining data that the contractor will supply to support integrated developmental testing (DT)/OT.
15. Establish warranty requirements, if applicable.	Lead for T&E works with the lead engineer and the KO on determining cost-effective warranty requirements, such as addressing and correcting defects (hardware, software, documentation) as part of the warranty.
16. Prepare an SSP and RFP (for competitive negotiated contracts).	Lead for T&E provides input to the SSP per the SOO/SOW, Section L (instructions, conditions, and notices to offerors or respondents), and Section M (evaluation factors for award) of the RFP.
17. Conduct source selection and award the contract to the successful offeror.	Lead for T&E participates on source selection teams.
18. Implement requirements for a contract administration office memorandum of	Lead for T&E provides input regarding the T&E support efforts for inclusion in the memorandum of agreement (MOA) and/or letter of delegation. The MOA should define product/system performance requirements and/or attributes.

Table 1-1 Acquisition Process Activities and the T&E Role

Typical Acquisition Process Activities	T&E Role
19. Monitor and control contract execution for compliance with all requirements.	PM, Lead for T&E, and program team perform programmatic and T&E monitor and control functions as defined in the contract. They assist the earned value management (EVM) implementation by monitoring the criteria for completion of T&E events, activities, and delivered products. They also use T&E performance criteria in the incentive/award plan.
20. Close out contract.	Contract close-out is mainly an accounting/ administration activity, but KO provides status updates to PM. Lead for T&E may have input regarding any T&E-related articles, such as M&S tools and final performance reports.
21. Document T&E lessons learned.	Lead for T&E and contractor partner should be capturing, and adjusting as necessary, lessons learned as the T&E effort progresses through the acquisition process. The lessons learned should be provided to the PM as part of the T&E close-out process and final PM report, as appropriate, to the program sponsor, or as directed.

1.7. Security Review, Public Release, and International Traffic in Arms Regulation (ITAR)

DoDI 5230.29, "Security and Policy Review of DoD Information for Public Release," implements policy in DoD Directive (DoDD) 5230.09, "Clearance of DoD Information for Public Release," and assigns responsibilities and prescribes procedures to carry out security and policy review of DoD information for public release. DoDD 5230.09 requires that a security and policy review be performed on all official DoD information intended for public release that pertains to military matters, national security issues, or subjects of significant concern to the Department of Defense. The program management office is typically the originator/owner of the weapon system classification guides. The PM, SE, KO, and Lead for T&E must work together to ensure documents have the proper control markings and that the public release process for T&E reports on the performance of contracted defense articles (at least in terms of unclassified export controlled test results) complete the security review process contained in DoDD 5230.09 and DoDI 5230.29.

The U.S. Government views the sale, export, and re-transfer of defense articles and defense services as an integral part of safeguarding U.S. national security and furthering U.S. foreign policy objectives. Authorizations to transfer defense articles and provide defense services, if applied judiciously, can help meet the legitimate needs of friendly countries, deter aggression, foster regional stability, and promote the peaceful resolution of disputes. The U.S., however, is cognizant of the potentially adverse consequences of indiscriminate arms transfers and, therefore, strictly regulates exports and re-exports of defense items and technologies to protect its national interests and those interests in peace and security of the broader international community. See Appendix A for additional information on Defense Export Controls and the International Traffic in Arms Regulation (ITAR).

2. PRE-SOLICITATION

The contents of this section will help you focus on and consider the most important contractual T&E items as you formulate the T&E strategy and approach. The discussion is applicable whether you are preparing for a weapons system acquisition program; command, control, communications, computers, intelligence, surveillance, and reconnaissance (C4ISR) system acquisition program; or automated information system acquisition program. A solid T&E strategy and approach foundation will facilitate the transition to the solicitation phase.

2.1. Planning

During the program life cycle, it is critical that the PM, SE, and T&E personnel recognize that early and consistent incorporation of T&E considerations and requirements begins at the onset of program planning during the materiel solutions analysis (MSA) and TD phases. The program acquisition strategy must be grounded in a technical approach with understandable, achievable, testable, and measurable performance requirements and reliability measures embodied in viable system solutions that are within cost and schedule constraints.

The PM and the program must be prepared to enter the EMD phase with cost, schedule, and expected system performance requirements balanced and synchronized. Important PM and team T&E considerations for possible use when beginning pre-solicitation activities are as follows:

- Ensure that program planning documentation, even in draft, such as the AS document or acquisition plan, AoA, TEMP, SSP, RMP, and the RFP are available, coordinated, and consistent. The SSP, RMP, and the resulting RFP should integrate the T&E policy directives and best practices from Government and industry.
- Ensure that the integrated T&E strategy and approach address the total life cycle of the program and consist of logically sequenced test events consistent with product or system development, demonstrated performance reviews, and satisfactory reliability metrics.
- Ensure that the specific test ranges/facilities and test support equipment are identified for each type of testing. Any shortfalls between the scope and content of planned testing with existing and programmed test range/facility capability must be identified with associated risk analysis. Ensure that any applicable requirements for OT&E are also addressed in addition to individual DT&E requirements.
- Incorporate T&E requirements in budgets and cost estimates in the program's T&E approach and achievable performance requirements, and the program's IMP, IMS, integrated master T&E schedule, and Earned Value Management System (EVMS). Program T&E cost and schedule realism must be supported by aggressive leadership, sound program planning, and timely application of resources along with execution of technical, T&E, and management processes. Ensure that operationally representative environments are available for system testing as well as OT to optimize integrated testing and efficiencies.

- Consider joint interoperability test command (JITC) interoperability and Net-Ready KPP certification. In addition, factor into the test strategy sufficient and early IA planning through the DoD guidance for DIACAP (Reference (q)) to ensure that operationally representative test environments and connectivity can be obtained.

2.2. Requirements

This guide addresses several kinds of requirements.

2.2.1. System Performance Requirements

The chief or lead SE is responsible for deriving system performance requirements. The Lead for T&E is responsible for ensuring that these requirements are testable and measurable. The approved performance requirements are the backbone of the T&E strategy, approach, execution, and reporting.

The system performance requirements should be performance-based, and potential system solutions must be based upon mature technology and lie within program cost and schedule constraints. These performance requirements are documented in the acquisition program baseline (APB) and should be in the SOO and based on the operational requirements stated in the ICD, or the follow-on CDD and associated JCIDS documentation. The preliminary system specification may include some of the JCIDS documents (or extracts from them), such as operational and system architectural views and CONOPs. The program office may also provide portions of the JCIDS documentation as reference material to aid the offeror's understanding of the operational requirements. The preliminary specification in the RFP is a precursor to the SPS that represents the program's functional baseline to be placed on contract. The functional baseline in the SPS is the first critical technical baseline established at the start of EMD.

2.2.2. Operational Requirements

Operational requirements are those written by the warfighter, operator, or user to express needed capabilities. Performance requirements, derived from operational requirements, must be established that correlate with program costs and schedule. If approach, execution, and reporting are not balanced at the start of the EMD phase or program award, the program has a high probability of incurring cost increases and suffering schedule delays or worse, a deficient system.

2.2.3. Reliability Requirements

In accordance with Reference (n), PMs formulate a comprehensive R&M program using an appropriate reliability growth strategy to improve R&M performance until R&M requirements are satisfied. The lead DoD Component and the PM, or equivalent, prepare a preliminary RAM and cost rationale report in support of the MS A decision. The TES and the TEMP should specify how reliability will be tested and evaluated during the associated acquisition phase. Reliability growth curves (RGCs) should reflect the reliability growth strategy and be employed to plan, illustrate, and report reliability growth. An RGC should be included in the SEP at MS A and updated in the TEMP beginning at MS B. The RGC will be stated in a series of intermediate goals and tracked through fully integrated, system-level T&E events until the reliability threshold is achieved. If a single curve is not adequate to describe overall system reliability, curves will be provided for critical subsystems with rationale for their selection. A method for identifying RGC risk is in Appendix F. It is highly recommended that reliability scorecards be included as part of proposals. Key for the T&E team is understanding all the stated and implied requirements and how to best test/evaluate those requirements. This understanding is based on sound analysis that uses integrated T&E, M&S, and a team composed of all stakeholders. The Lead for T&E ensures the T&E strategy and approach addresses system of systems (SoS) and joint T&E to the extent necessary to adequately demonstrate performance in the expected operational environment with realistic T&E events and schedule. The Lead for T&E along with test team members should develop a requirements/testability crosswalk matrix depicting how each requirement will be tested. Use the evaluation framework found in the TEMP as the basis for the matrix. See Chapter 9 of Reference (b).

2.3. TES and the AS Document/Acquisition Plan

The PM and Lead for T&E must recognize and emphasize the importance of a sound T&E strategy and approach to the program. The recognition begins with the statement of required capability, resulting in an approved system definition that provides a product meeting the user's needs. There is no "one size fits all" approach for programs, but disciplined adherence to proven T&E processes and practices will lead to a sound T&E strategy and approach. When developing the T&E strategy and approach, consider that the single most important step for avoiding suitability failures is to ensure that programs are formulated to execute a viable systems engineering and T&E strategy from the beginning, including a robust RAM program that includes design for reliability (DfR) and for reliability growth and development.

The Government TES and TEMP are the foundation T&E documents supporting the acquisition strategy and the PM's program schedule, and contain key items to consider when developing the SOO, SOW, PWS, and RFP. The Government's T&E strategy and approach should describe what is to be accomplished. The offeror's integrated T&E approach provided in the proposal will expand on how the offeror intends to execute the integrated T&E program, applying its domain experience and corporate best practices. The Government TES, and then TEMP, should be prepared as early as possible to properly influence the acquisition process by providing a carefully planned T&E strategy and approach to meet the programmatic and operational needs.

This strategy and approach become very important if the acquisition strategy and engineering strategy employ incremental development and fielding. TES/TEMP development should begin in parallel with the analysis of operational requirements so the T&E strategy and approach are consistent with the required capability. The Government should share the draft TEMP and the draft preliminary system specification with industry representatives to obtain their perspectives on the T&E strategy and approach. In addition to the TES/TEMP, the program requires supporting documents such as the SEP, AS, RPP, SSP, and ICD/CDD. These program documents capture information important to developing the T&E strategy and approach.

2.3.1. Working With Industry

During the pre-solicitation phase of a program, the T&E process should be applied to set the stage for future expectations. The Government is in the leadership role in this stage, and early industry input can provide critically important insights into the technical and performance challenges, program technical approach, and key business motivations. Lessons learned from past programs suggest the pre-solicitation process can be very productive when a highly collaborative environment is created, involving the user, acquisition community, and industry personnel. The program should ensure early and frequent industry involvement while developing the T&E strategy and approach and the formulation and development of the system performance requirements. Industry can provide important insight into the T&E and business aspects of the program. The Government should include its T&E strategy and approach in the draft RFP to foster this synergism and interaction. Notwithstanding the desire to work with industry and obtain insight on T&E solutions from potential contractors, Government personnel should be aware that individual contractors will have potential biases that will intrude into their recommendations. The Government, therefore, should seek independent counsel from numerous sources to minimize the impact of any specific contractor's potentially biased recommendations.

2.3.2. Formula-Type Incentives and Award Fees

There are two broad types of incentive contracts: (1) those that rely on the application of predetermined, formula-type (objective) incentives; and (2) award-fee contracts, in which the award amount is determined by the Government's (subjective) evaluation of the contractor's performance.

Both types of incentive contracts are designed to achieve specific acquisition objectives by establishing reasonable and attainable targets that are clearly communicated to the contractor, including appropriate incentive arrangements designed to motivate contractor efforts that might not otherwise be emphasized and discourage contractor inefficiency and waste. Most incentive contracts include only cost incentives, which take the form of a profit or fee adjustment formula and are intended to motivate the contractor to effectively manage costs. See Subpart 14.4 of Reference (e). No incentive contract may provide for other incentives without also providing a cost incentive or constraint.

In developing appropriate incentives, the Government must take care to provide incentives only for the desired behavior, not for actions that are counterproductive or for requirements that the contractor would otherwise be obligated to perform. Incentive increases or decreases are applied to performance targets rather than minimum performance requirements. Incentives are directly linked to expectation setting, understanding, and interactive management. Incentives and motivations must support the overall program needs and not weaken a specific aspect of the program. If the contractor develops an internal test capability for a capability that already exists within the Major Range and Test Facility Base (MRTFB), there must be clear evidence that it is in the best interest of the Government and program by conducting a cost-benefit analysis (CBA).

2.3.2.1. Formula-Type Incentives

Formula-type incentives are based on a single criterion or multiple criteria that can be objectively measured. The Department is increasingly moving toward incentives based on objective criteria, according to Reference (r): "It is the policy of the Department that objective criteria will be utilized, whenever possible, to measure contract performance." Also see the memorandum from the Under Secretary of Defense for Acquisition, Technology, and Logistics Memorandum (Reference (s)).

For example, a cost incentive would be that the additional cost for every dollar over the target cost of the contract would be split between the Government and the contractor based on a fee adjustment formula (i.e., share ratio). Including incentives for T&E excellence in addition to the cost incentive can be an important aspect of the program acquisition strategy and should be an explicit consideration for any development or test program contract. The incentive strategy must be balanced with the program cost, schedule, and performance requirements reflected in the program documentation. Incentives reinforce the Government's emphasis on T&E leadership, planning, and execution with the contractors. Incentives beyond the required cost incentive may be monetary, nonmonetary, positive, or negative, but regardless of their structure, the goal is to encourage high-quality performance to achieve program goals. The PM must prepare an incentive fee determination plan to document the process that will be used to determine the incentive fee.

Incentives for motivating excellence in the T&E portion of a program may be based on schedule or performance, but an incentive contract cannot provide for other incentives without also providing a cost incentive or constraint (Subpart 16.402 of Reference (e)). Some of the T&E criteria are inherently mixed with other criteria, especially technical criteria, including risk management, timely data delivery, and access. Incentives should be tied to specific test events, such as demonstrating a specific capability in the system integration laboratory or testing a critical capability with a full-scale test article.

The incentives applicable to T&E have tended to be subjective, award-fee measures, which will be discussed in section 2.3.2.2. When structuring incentives for the entire program, the RFP team must keep in mind the Federal Government's policy to not incentivize minimum performance requirements and to avoid the potential dangers of incentive dilution, incentive contradiction, and unintended adverse consequences. For example, small increases in incentivized performance may have undesirable impacts on other program elements that are important but not incentivized. Or, a contractor's desire to earn schedule incentives could detract from sound engineering decisions. Schedule-based incentives may diminish the intended benefits of the test activity (e.g., the data collected cannot support integrated testing strategy).

The incentives should take into account non-test items that could affect the length or productivity of the test program. For example, if a radar system is not ready for testing at the same time as the rest of the weapon system, the test program could be delayed or lose efficiency because the program has to repeat test events when the radar is installed. In that case, an incentive placed on delivery of critical subsystems to the test program would have a greater effect on test program efficiency than any incentive applied directly to the test program itself. However, this also may be accomplished through a modification in delivery schedules of the critical subsystems. In general, focus incentives on demonstrating that key programmatic and technical risks are resolved as soon as possible, and avoid any incentives that may drive the contractor to delay testing inappropriately.

The contractor can be incentivized to use preexisting Government ranges, facilities, and instrumentation that is sized, operated, and maintained to provide T&E information to Government T&E users. If the contractor develops internal T&E capabilities that duplicate existing Government facilities, the acquisition program may incur additional costs that could be avoided. Incentives can also be tied to the contractor using preexisting Government test ranges/facilities to include instrumentation. As a national asset, the MRTFB is sized, operated, and maintained to provide T&E information to DoD Component T&E users in support of DoD research, development, T&E, and acquisition processes. If the contractor develops an internal test capability for a system that already exists within the MRTFB, a cost penalty will be incurred.

- **Use of Government Test Facilities**. The Lead for T&E will take full advantage of existing investments in DoD ranges, facilities, and other resources, including the use of embedded instrumentation. Test teams should plan to use MRTFB facilities and capabilities first, followed by Service test facilities and capabilities, followed by non-DoD Government facilities.
- **Use of Nongovernment Facilities**. Contractor facilities should be used only when Government facilities are not available, cannot be modified, are too expensive, or are impractical to use. If the strategy for T&E calls for testing at nongovernment facilities, the PM must conduct a CBA, include these facility requirements in the RFP, and document the final choice in the TEMP.

2.3.2.2. Award Fees

The application of award fee incentives is generally associated with cost-reimbursement contracts but may be used in either fixed-price or cost-reimbursement contracts. An award fee provision may be used when the Government wishes to motivate a contractor and other incentives cannot be used because contractor performance cannot be measured objectively (See Subpart 16.404 and 16.405-2 of Reference (e)). The award fee approach is suitable for use when it is neither feasible nor effective to devise predetermined objective incentive targets applicable to cost, technical performance, or schedule.

Although award fee incentives can produce positive effects, the effort required for periodic evaluations in accordance with the award fee plan (e.g., continuous monitoring, midterm analyses, final analyses, and periodic reports) must also be considered, particularly for smaller program teams. Consider the investment in resources versus incentive gain before deciding to use an award fee approach. Award fee criteria need specific data and performance examples to make an award fee determination. As subjective measures are used, the contractor must clearly understand expectations and be promptly advised of any problems or issues that may affect the award determination.

The contractor earns the incentive awards through a subjective evaluation process conducted by an Award Fee Review Board described in an award fee plan. For example, if the program requires the contractor to develop a test bed, the award fee incentive could be related to the test bed development, test, and acceptance according to the schedule, cost, and test bed performance requirements. This incentive approach allows the Government to reward exceptional contractor performance while considering the conditions under which it was achieved, normally in such areas as quality, timeliness, technical progress, technical ingenuity, and cost-effective management. The Government should avoid making early completion of technical reviews an award fee criterion because such an incentive could discourage the conduct of sufficiently thorough event-based reviews and therefore be counterproductive. Appendix C lists sample T&E award fee criteria. Table 2-1 lists 14 items to consider when developing T&E award fee criteria.

Figure 2-1 T&E Award Fee Considerations

1. Contractor has executed the T&E strategy and approach in accordance with the TES/TEMP/test plan, and integrates management plans/tools.
2. Contractor has implemented and demonstrated a disciplined T&E management process to capture test entrance, exit, and success criteria with clearly defined metrics.
3. Contractor has presented a well-thought-out trade study and/or limited development test (LDT) plans for the program and provides evidence of systematically evaluating all aspects of the system. The trade studies utilize common sets of critical trade parameters that are focused on the critical performance, schedule, and cost requirements of the program. Trade studies are documented and archived to establish an audit trail for the principal technical decisions on the program. The contractor conducts LDTs to test and evaluate specific critical aspects of system performance.
4. Contractor has demonstrated that T&E data ownership, control, access, sharing, completeness and accuracy, and delivery support the T&E strategy and approach.
5. Contractor continually demonstrates timely and efficient preparation of T&E plans and reports as the system is progressively described to its lowest level of detail.
6. Contractor uses M&S to minimize the number of tests, which results in overall lower costs. M&S must undergo verification, validation, and accreditation (VV&A).
7. Contractor has implemented a process to track test failures, analyze and establish corrective actions, and provide feedback into plans and procedures to improve T&E efficiency. The contractor has a deficiency reporting (DR) system that is compatible with and feeds into the Government-run DR system.
8. Contractor has established and implemented an event-based T&E process through the use of technical performance measures (TPMs) to include reviewing events with entry, exit, and success criteria.
9. Contractor demonstrates effective risk management, actively involving the Government to assess major risk areas, and establishes specific risk mitigation plans that are integrated into program plans.

Figure 2-1 T&E Award Fee Considerations

10. Contractor flows T&E processes and plans to the subcontractors and actively involves the subcontractor team in T&E baseline management, configuration management, requirements management, and risk management activities.
11. Contractor has a disciplined action item tracking system that documents system and subsystem, if applicable, performance problems/issues that require program management attention.
12. Contractor has an exceptional record in meeting milestones and due dates and effectively uses T&E metrics to manage the T&E program.
13. Contractor has implemented Department-level policy and guidance, including JCIDS planning processes, scientifically based test design (i.e., design of experiments (DOE), HSI testing, and testing in a joint environment.
14. Contractor has implemented opportunities for integrating contractor testing, DT, OT, interoperability, security, IA, and DIACAP certification with the goal of developing cost-effective test programs with shorter schedules.

2.3.2.3. Information on Incentives

Part 16 of Reference (e), the DFARS (Reference (d), Service FAR supplements, and individual incentive and award fee guides (e.g., Air Force Award Fee Guide (Reference (i)), Air Force Guide Award Term/Incentive Options (Reference (t)), Army Award Fee Guide (Reference (j)), and the Navy-Marine Corps Award Fee Guide (Reference (k))) provide additional information, address ways to structure incentive and award fee plans, and provide examples. Other applicable references and guides include a memorandum from the Office of the Deputy Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD(AT&L)) Memorandum on Award Fee Contracts (Reference (u)), DAU's Acquisition Community Connection (Reference (v)), and the Under Secretary of Defense for Acquisition and Technology Memorandum, "Incentive Strategies for Defense Acquisitions" (Reference (w)), which provide details on different incentive approaches.

2.4. Market Research

FAR Part 10 (Reference (e)) requires the Government's acquisition strategy to include the results of market research. FAR Part 10 implements the requirements in sections 253a(a)(1) and 264b of title 41 and section 2377 of title 10 of Reference (h). Market research is one method to establish the availability of products and the suitability of commercial products (e.g., COTS products) to meet the potential Government system performance needs. Such research supports the acquisition planning and decision process by supplying technical and business information about commercial and DoD technology, products, and industrial capabilities.

Market research is used to obtain current information on companies' maturity model level rating and their application of rated processes within specific domains of their company. The maturity model rating is not the sole determinant of process maturity. The corporate commitment to continuous process improvement (CPI) with documented plans and maturity milestones is also an important element. Frequently during the pre-solicitation and RFP preparation phase of a program, the Government team seeks business, T&E, and acquisition planning information via a request for information (RFI). The Government usually sends these requests via the Government-wide point of entry found at the Federal Business Opportunities (FedBizOpps) Website (Reference (x)). The RFIs solicit data from interested industry sources, but such responses might be limited because the request for data and information is unfunded. The RFI can be used to supplement market research and to secure specific types of T&E data, including the extent of the companies T&E domain experience and details on their T&E best practices. RFIs can provide valuable insight on how potential offerors have integrated their technical, T&E, and management processes to effectively manage prior programs. Each year, the MRTFB activities are required to submit a notice, via FedBizOpps, that describes the nature of the anticipated commercial work and invites private sector responses proposing capability to perform these T&E services.

2.5. Industry Days

Before release of a formal RFP, the Government may hold Industry Days to inform industry about the technical requirements and acquisition and T&E strategies, and to solicit industry input for the pending program. During this time, communications are unencumbered by the formality and limitations associated with the formal RFP/source selection process. T&E personnel need to avail themselves of this opportunity for free and open communications. They should emphasize the importance of the significant aspects of T&E requirements (such as M&S, hardware-software and system component integration T&E, use of test beds, prototypes, incremental T&E and fielding, having interoperability architectures, and identification of specific ranges) to resolve T&E complexities and mitigate actual or anticipated program risks. The Government should initiate discussions of the following seven T&E topics during Industry Days discussions.

1) **T&E Strategy and Approach**. Continually emphasize the importance of the overall technical approach and associated T&E strategy and approach. The Government-prepared TES/TEMP should be made available to industry, in accordance with DoD Component direction and guidance.

2) **M&S Users**. Discuss M&S testing (especially the VV&A process and proprietary rights) and any trade studies, LDTs, and analyses that have been conducted during the requirements generation process. Although solution alternatives are studied during this phase of the program, the emphasis should remain on the resulting performance requirements and not on the specifics of the alternatives. Government trade studies, LDTs, and analyses should be made available to industry as appropriate.

3) **Potential T&E Solutions**. Although it is necessary to investigate potential T&E solutions that are responsive to the requirements, the Government team should avoid becoming fixated with the solutions. The user sometimes becomes enamored with what he or she likes, the acquisition team focuses on the solution that works, and industry has a solution it wants to sell. Instead, the team should focus on establishing the cost-effective T&E processes and events that generate appropriate

technical and operational data to allow decision makers to make informed decisions.

4) **Supporting Management Processes**. T&E members need to emphasize that potential offerors must have T&E management processes to implement during program execution. The Government team should have a clear understanding of system/subsystem requirements, encourage the offerors to discuss their T&E approach, and encourage the potential offerors to document their approach.

5) **T&E Approach**. T&E members need to address the T&E approach and how it was established. This is an excellent opportunity to reinforce the importance of the T&E processes and schedule for the program and for the Government to describe its T&E approach to the program.

6) **Corporate Proprietary Information**. Keep in mind that prospective offerors exercise extreme caution during open sessions for fear of compromising a competitive advantage or revealing a perceived weakness. During one-on-one sessions, the discussions are more open, but be careful to provide all offerors with equivalent information about the Government's needs without divulging potential solutions considered by other offerors.

7) **Areas of Mutual Interest**. Identify areas of interest and encourage prospective offerors to provide data, insights, and suggestions that facilitate the transition into EMD with sound performance requirements and a well-structured T&E approach. The agenda and topics should not be left solely to the discretion of the offerors.

For additional information on exchanges with industry before receipt of proposals, see the other eight techniques discussed in Subpart 15.201(c) of Reference (e).

2.6. Division of Responsibilities/Authority

Additional Government team considerations for working with industry are the division of responsibilities between the Government and the contractor, the definition of contractor testing and Government testing, and the level of authority granted to each to execute the test program. The contract should be clear on what the contractor is expected to deliver in terms of articles, data, performance, or services. However, T&E programs usually involve a shared responsibility in the planning, execution, and reporting of T&E. If this shared responsibility and authority are not clearly addressed during contract formulation and award, then any misunderstandings will cause problems during program execution. The problems can range from minor discussions over who can approve test plans, to major disconnects, such as missing equipment, that can bring the program to a halt.

The strategy for planning and executing the test program needs to be agreed upon prior to release of the solicitation. One strategy consideration concerns overall control of the test program: *Will the contractor run everything with the Government testers in a support role at the contractor's facility? Will it be shared? Or, will the Government testers at Government ranges/facilities be in control with the contractors in a supporting role?* Remember, for OT, the contractor can be involved only to the extent that it will be involved once the system is fielded. Responsibilities related to planning detailed tests and controlling execution of test events also need to be considered. In addition, responsibilities for conducting test-related safety analyses and mitigating test risks must be considered during SOW and RFP generation. Some of the answers will be driven by the choice of test ranges and facilities to be used (e.g., contractor or Government), but such issues still must be explicitly considered.

Another factor in addressing the level of responsibility of the contractor versus the Government is the overall level of system performance responsibility assigned to the contractor through the contract. *Will the contractor have total system performance responsibility and be expected to handle all of the integration issues for the total system and deliver end-system performance? Or will the contractor be responsible for only one element of the total system, with the Government or another contractor becoming the system integrator and accepting the risks associated with delivering end-system performance?* The choice will affect the way in which the Government works with the contractor and the division of responsibilities and authority between the Government and the contractor.

2.7. Request for Proposals (RFPs)

The RFP is a solicitation used in negotiated acquisition to communicate Government requirements to the prospective offerors and to solicit proposals. Subpart 15.204 of Reference (e) specifies that the format and content of RFPs and contracts be prepared in accordance with specific guidelines called the Uniform Contract Format (see Figure 2-2).

Part I – Schedule
- A – Solicitation/contract form
- B – Supplies or services and process/costs
- C – Description/specifications/statement of work
- D – Packaging and marking
- E – Inspection and acceptance
- F – Deliveries of performance
- G – Contract administration data
- H – Special contract requirements

Part II – Contract Clauses
- I – Contract clauses

Part III – List of Documents, Exhibits, and Other Attachments
- J – List of Attachments

Part IV – Representations and Instructions
- K – Representations, certifications, and other statements of offerors or respondents
- L – Instructions, conditions, and notices to offerors or respondents
- M – Evaluation factors for award

Figure 2-2 Uniform Contract Format

The RFP typically includes two kinds of documentation: program documents and RFP documents. Figure 2-3 depicts the flow from program documentation to populate typical RFP sections to a typical proposal.

- **Program Documents**. The AS, program IMP or top-level program roadmap, incentive plan or award fee plan, Government SEP, TEMP, and the preliminary SPS are the program's important documents that are typically attached or referenced in the RFP and may be included in an "Offeror's Library." These documents describe the Government's management, technical, and T&E approach to the system acquisition along with the required system performance requirements and other important program planning elements.

- **RFP Documents**. A typical RFP includes a model contract with any special contract requirements, contract line item numbers (CLINs), SOO, SOW, CDRL, WBS, evaluation criteria (Section M), and instructions to offerors (Section L). The RFP (in concert with the program documents) defines the program to be proposed.

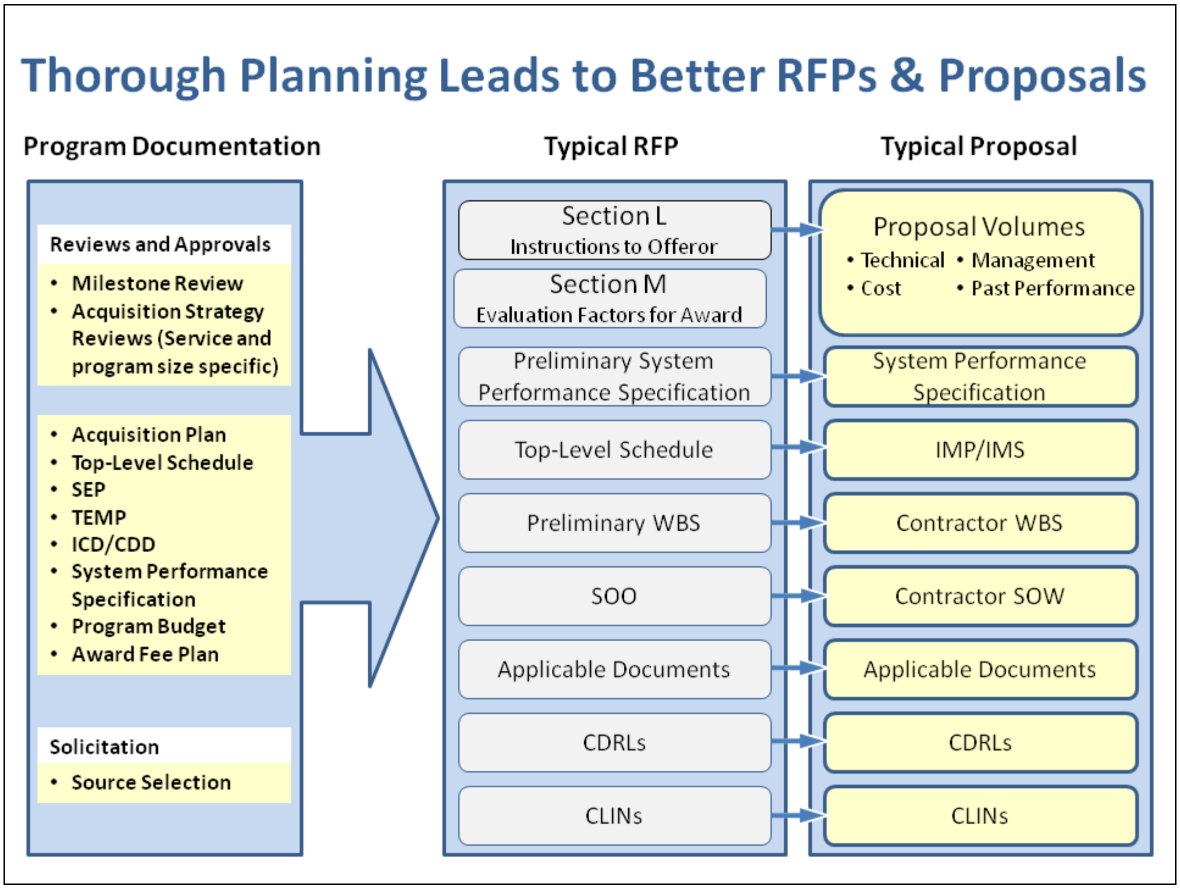


Figure 2-3 Relationship of Program Planning to a Typical RFP and Proposal

Early preparation of the Government TEMP is an important step to foster synergy among RFP sections. An integrated approach, developed specifically for each program, will result in a high degree of synergism and integration of all RFP and proposal elements. For instance, the SOW, PWS, IMP, IMS, SEP, TEMP, model contract, and the critical processes are all interrelated. The following subsections discuss the core RFP documents that contain substantive T&E material and the applicable companion proposal documents. Sections C, L, and M are the primary parts of the RFP influenced by the T&E approach to the program.

The RFP captures and amplifies the acquisition, technical, T&E, and support program strategy. There is a natural flow of information from the program strategy, to RFP, to proposal, and the resulting contract. Each program must develop the RFP according to the program strategy. Some items are required for source selection purposes only, such as the proposal volumes and/or past performance information. Some items will become parts of the contract, such as the IMP, SOW, and SPS.

2.7.1. Statement of Objectives (SOO)

The SOO is that portion of a contract that may establish a broad description of the Government's required performance objectives. The SOO delineates the program objectives and the overall program approach. The SOO, along with the preliminary SPS (covering the technical performance requirements), provides offerors with guidance for proposing a program to meet the user's needs. The SOO is an RFP document that does not become part of the ensuing contract.

A PWS may be prepared by the Government or result from a SOO prepared by the Government in which the offeror proposes the PWS in accordance with Subpart 37.602 of Reference (e). The SOO is replaced by the PWS; the PWS becomes part of the contract.

The Government desires an efficient and integrated experimental design and analysis (i.e., DOE) approach linking the contractor's design process, proposed M&S efforts, hardware-in-the-loop (HWIL) utilization, and any planned ground or flight test associated with this effort. Selected test points should be tested in the HWIL as part of formal qualification testing. The Government expects efficient and effective tests with statistically significant results over a broad range of operational conditions. Test plans should cite the statistical risks implied by the proposed test programs. Specifically, tests should be designed to obtain data that: (1) support design space trades as part of the development process, (2) assess performance, and (3) predict performance over the system's operational space. The Government expects that the contractor will implement this experimental design and analysis approach to take advantage of the greater understanding of product physics, identify sensitivities to the parameters of interest, and utilize relevant statistics to reduce schedule and costs while lowering overall program risk.

Section C of the RFP contains the detailed description of the products to be delivered or the work to be performed under the contract and the preliminary SPS. The preliminary SPS is addressed in section 3.13.1 of this guide. Other contract requirements documents may be included such as sample IMP event descriptions, CDRL, Contract Security Classification Specification (DD Form 254), and pricing matrices. Table 2-2 contains text for inclusion in a SOO that emphasizes the main T&E themes of this guide. Specific program requirements and the program strategy are used to modify this example.

Table 2-2 T&E Content for the SOO

Statement of Objectives
The T&E approach will capitalize on best practices from industry domain experience and will implement DoD T&E policies. The program shall: 1. Document the T&E approach in an integrated Government TEMP that covers the life of the program. Reference (a) defines the life of the program. 2. Utilize contractor T&E best practices and processes to reduce cost. Includes agile and mature technical and management program processes based on company processes that undergo continuous improvement throughout the program's life cycle. Policies and processes shall flow down to the lowest level of the contractor (subcontractors, teammates, or vendors) team. 3. Implement event-based program milestones (e.g., critical design review (CDR)) and integrated schedules (e.g., integrated master T&E schedule). Implement event-based T&E events and reviews involving Government and industry SMEs. 4. Use contractor configuration management processes to control the configuration of the T&E data in a common T&E database. Provide real-time access to the T&E baseline data for program participants. 5. Enhance opportunities for incorporation of improved capabilities and advanced technology using the modular open systems approach (MOSA). Encourage use of commercial products/processes/standards. 6. Include Government participation on integrated product teams (IPTs)* to gain insight into program progress. 7. Ensure that the requirement for a RPP is documented. 8. Implement a comprehensive risk management process that also includes risks associated with the program's critical path, to systematically identify and eliminate/mitigate cost, schedule, technical, and performance risks. 9. Institute a requirements management process coupled with a T&E baseline management strategy that supports the TD and EMD phases, as applicable, and an orderly transition to the production, deployment, operation, and support acquisition phases. 10. Ensure that the contractor has an efficient and integrated experimental design and analysis approach (i.e., DOE) that supports program execution. 11. Require contractor participation as appropriate in Government reviews in which T&E matters are discussed (e.g., certification of test readiness, technical reviews, and data reviews).

* T&E SMEs may participate in different teaming arrangements, including T&E IPTs, T&E WIPTs, and program-specific teams such as contractor/combined test teams (CTTs), a combined T&E task force (CTF), or ITTs. The title by itself is not important. The key to a team structure is the charter, which lists the roles, responsibilities, products, and stakeholder membership.

2.7.2. SOW/PWS

The SOW/PWS is that portion of a contract that establishes and defines all non-specification requirements for a contractor's efforts, either directly or with the use of specific cited documents. The offeror may provide a SOW to be included in the negotiated contract. The Government may provide a SOW/PWS as part of the RFP instead of a SOO, in which case the offerors will tailor the SOW/PWS in their proposals depending on their specific solutions to the requirement. The SOW/PWS should accomplish the following:

- Describe the T&E events and activities to be accomplished that reflect the T&E approach to the program as described in the TEMP.
- Reflect use of T&E processes across the program that are critical for program success. Processes include reliability growth planning, technology maturity assessment, management of performance deviations and waivers, performance baseline control, risk management, configuration, experimental design and analysis approach (i.e., DOE), integrated testing, and T&E data management, including Government access and sharing of contractor data, tests, and results.
- Plan for and support T&E events and event-based reviews as defined in the TEMP or the program plan.
- Address the T&E baseline management process, associated T&E data, and Government-approved stakeholder access to all T&E data, including M&S data.
- Provide for TEMP updates and CPI consistent with corporate improvements, technical changes, and program needs.
- If a Government SOO has been developed, include a cross-reference matrix tracking the Government SOO requirements to the proposed SOW. The SOW should be structured for the proposed system solution and not restricted by the structure of the Government's SOO.
- Include the necessary contract language to ensure that an RPP is delivered.
- Address the following items, as necessary, relative to the T&E strategy and approach: contractor test plan, detailed test plans and reports, T&E support for Government-conducted tests, test instrumentation, TRRs, failure review boards, DR, and T&E WIPT support.

The contractor SOW/PWS addresses the requirements in the SOO or RFP SOW, other sections of the RFP, and derived requirements based on the offeror's approach. The SOW should include those T&E tasks and activities that the contractor is required to execute during the contract. The T&E approach relies heavily on contractor's processes and practices, and the SOW should address the application of these processes and practices during DT&E, OT&E, and sustainment as applicable to the program. It is generally not the intent to put the specifics of the contractor's individual processes and practices on contract, but the SOW should recognize the application of key T&E processes and practices on the program. The SOW should address the Government's requirement and not a contractor's solution. When a contractor proposes a detailed SOW, it must still be stated in terms that describe the Government's requirements. Table 2-3 provides a sample content for the SOW/PWS.

Table 2-3 T&E Content for the SOW/PWS

Sample Instruction for Proposing T&E Activities in a SOW/PWS
The offeror shall provide a SOW/PWS to be included in the negotiated contract. The SOW/PWS shall: 1. Describe the T&E work/tasks/activities to be accomplished on the program that reflect the T&E approach to the program as described in the TES/TEMP. 2. Identify the role of M&S to be used in support of the T&E process and the documented VV&A of any M&S to be used. 3. Reflect use of T&E processes across the program that are critical for program success (e.g., requirements management, performance baseline control, risk management, configuration and data management, and interface management). 4. Provide for event-based reviews as defined in the integrated master T&E schedule and/or the program master schedule. 5. Address the T&E baseline management process, associated data, and stakeholder access to all T&E data, especially the handling and accountability of expected performance deviations or waivers. 6. Provide for TES/TEMP updates and CPI consistent with corporate improvements and program needs. 7. Include a cross-reference matrix showing the tracking of Government SOO or SOW requirements in the proposed SOW. The SOW should be structured for the proposed system solution and not restricted by the structure of the Government's SOO or SOW. 8. Provide the proposed RPP format and content. 9. Describe the closed-loop Failure Reporting and Corrective Action System (FRACAS) strategy in terms of methodology, processes, and database(s) used to support the contract and throughout the system life cycle. The proposed contractor DR database must be compatible with (i.e., feed into) the Government's DR database. 10. Provide personnel and documentation (drawings, etc.) support to Service OTA during operational assessments and test events.

2.7.3. RFP T&E Insertion

The following information will assist the T&E Lead when working with the PM and the IPT in the development of the RFP and helps ensure the program RFP from a DT&E perspective tracks with program acquisition documents along with requirements prior to release to industry.

- **T&E Management.** The T&E Lead should ensure that the RFP describes overall T&E management structure, responsibilities, experience of T&E staff, and application of T&E best practices.

- **T&E Data.** The T&E Lead should ensure that the RFP describes the contractor's approach to technical data, to include management, control, access, and delivery of T&E data.

- **M&S.** Ensure (if applicable to program) that the RFP describes allocation of M&S responsibilities, expectations, and M&S tools.

- **RAM.** Ensure that the RFP describes the approach and procedures to perform RAM.

- **IA.** Ensure (if applicable to program) that the RFP describes the contractor's IA responsibilities.

- **T&E Planning and Resources.** Ensure that the RFP describes the change management process for updates to test plans and test assets. Describe Government and contractor test resources required. Ensure that a business case analysis was conducted and documented within the TEMP for use of contractor-unique resources vice Government-owned facilities.

- **Software.** Ensure (if applicable to program) that the RFP describes the responsibilities of the contractor and Government during test execution. Ensure that the process for contractor DR and resolution is described.

The following documents will assist in the development of the RFP: Draft RFP (with CDRLs), SSP, Program AS document/Acquisition Plan, TEMP, and any requirement documents (ICD, CDD, or CPD). The T&E Lead should see Subpart 15.204-2 and Part 16 of Reference (e) for additional clarification and/or guidance.

2.8. T&E Focus Areas

There are 10 T&E interest areas the PM team should address in the planning stage, prior to issuing a solicitation for a contract: integrated testing, shared test data access, system of systems (SoS), test assets, ranges and resources, reliability, modeling and simulation (M&S), government furnished equipment (GFE), safety, and software.

2.8.1. Integrated Testing

Integrated testing is defined in Reference (y) as "…the collaborative planning and collaborative execution of test phases and events to provide shared data in support of independent analysis, evaluation and reporting by all stakeholders particularly the developmental (both contractor and government) and operational test and evaluation communities." The PM and Lead for T&E need to consider the availability of in-house or DoD Component T&E resources, as well as contractor use, relationship, and responsibilities for DT&E, OT&E, interoperability, IA, security, and other equivalent types of T&E activities, to include DIACAP C&A. The PM and Lead for T&E need to consider such questions as:

- *Who will be in charge of the testing – Government or contractor?*

- *Will Government personnel "work" for the contractor (i.e., Government-furnished personnel)?*
- *Who is accountable for test conduct and reporting?*
- *What is the Government's T&E oversight role and process?*
- *Will the Government witness the testing at the contractor's facility?*
- *Will the Government receive all pertinent contractor raw test data?*

The contractor and Government's T&E roles and responsibilities must be clearly, accurately, and completely identified. Subpart 9.5 of Reference (e) provides the responsibilities, general rules, and procedures for identifying, evaluating, and resolving organizational conflicts of interest. Specific statutory and regulatory guidance exists with respect to contractor involvement in OT&E and LFT&E. DoD Components have specific guidance relative to contractor involvement in their respective acquisition programs.

System contractors may be beneficial in providing logistic support, test failure analyses, and software and instrumentation support that could increase the value of unprocessed OT&E data. Clear explanations of how system contractor support will be used and the mitigation of possible adverse effects must be described in the TEMP and OT&E plans to ensure no violation of the prohibitions in section 2399 of Reference (h).

2.8.2. Shared Test Data Access

Most systems will utilize technology and subsystems developed for other programs or in prior efforts. To take advantage of this prior data, and data generated during contractor development, the issue of data access needs to be addressed. Resolving the issue may touch on data rights issues, which can be a source of contention. The data access issue does not automatically mean buying all the data packages from the contractor. Instead, it means ensuring that the Government will have access to the needed data in the future. Perhaps the best outcome that can be negotiated in the contract is the fee or rate to be paid for whatever data is needed in the future.

Negotiating the data access issue early, during the competitive portion of the contracting process, will minimize the cost for the data requested later during the execution of the contract. Note that data access could be considered from both perspectives; the contractor may want access to data the Government has or is aware of concerning technologies that the contractor needs. Typically, if contractor test data is to be used as part of the independent system evaluation, the Government will require that the test be witnessed by the tester, evaluator, or PM. Data access also means that contractors are authorized to use the data, for example, information technology 1 or 2 or 3 access permissions, and that the contractors possess the required security clearance.

Testing requirements in the RFP should include procedures for ensuring the pedigree of the data. This should include that government will have review and approval of contractor test plans prior to execution of test event, government witness of test event, and government review and approval of final test report/analysis.

2.8.3. System of Systems (SoS)

Expected product/system interoperability should be clearly identified in the SOO and CONOPs and will drive the T&E strategy, needed resources, and schedule. For example, does the product/system being developed stand alone, or is it part of an SoS? What is the relationship between this system and the other systems? Are the boundaries/interfaces between systems well defined?

2.8.4. Test Assets

A significant costing topic is the number of test assets required for conducting the necessary test cycles during DT, OT, LFT&E, IA, security, interoperability, DIACAP C&A, and contractor testing. The number of test assets required for conducting DT&E, OT&E, IOT&E, and LFT&E is typically recommended by the T&E WIPT with Director of Operational Test and Evaluation (DOT&E) concurrence and documented in the TEMP that is approved by the Office of the Secretary of Defense. These determinations should include identification of spares. Consideration of this topic must be in conjunction with M&S expectations, any statutory and/or regulatory requirements, and required sample size, as determined by experimental design and analysis (such as DOE) or equivalent analysis, necessary to support the stated performance confidence levels.

2.8.5. Ranges and Resources

The identification of test ranges, facilities, and other needed resources (such as personnel, equipment, and test organizations for DT&E, OT&E, and LFT&E) should not be delayed until the final stages of TEMP approval. The test ranges, range resources, equipment, and personnel should be identified to the extent possible in the T&E strategy development process. Especially, it must be clear which DoD assets the Government requires the contractor to use or the contractor should specifically identify and justify use of its own test resources. Government and contractor test facilities should be compared to ensure that there is no duplication and that the most appropriate facility to conduct the T&E is identified. If Government test facilities are required, ensure that the contract with the DoD contract sponsor provides the use of test support from the Government T&E facility or capability at the established rate in accordance with chapter 12 in DoDD 7000.14-R, volume 11A. Otherwise, defense contractors will be charged as commercial customers.

2.8.6. Reliability

The offeror is expected to develop and provide an RPP to achieve the following four objectives: (1) understand the Government's requirements, (2) design product/system for reliability, (3) produce reliable products/systems, and (4) monitor and assess user reliability.

The RPP should accomplish the following:

- Provide visibility into the management and organizational structure of those responsible and accountable (both offeror and customer) for the conduct of reliability activities over the entire life cycle.

- Define all resources required to fully implement the reliability program.
- Include a coordinated schedule for conducting all reliability activities throughout the system life cycle.
- Include detailed descriptions of all reliability activities, functions, documentation, processes, and strategies required to ensure system reliability maturation and management throughout the system life cycle.
- Document the procedures for verifying the implementation of planned activities and for reviewing and comparing implementation status and outcomes.
- Manage potential reliability risks due, for example, to new technologies or testing approaches.
- Flow reliability allocations and appropriate inputs (e.g., operational and environmental loads) down to subcontractors and suppliers.
- Include contingency-planning criteria and decision making for altering plans and intensifying reliability improvement efforts.

The RPP is expected, at a minimum, to address the following 12 reliability topics. Specific descriptions of each of the activities may be found at Appendix B and the DAU Website at https://acc.dau.mil/CommunityBrowser. An example of a reliability scorecard can be obtained by sending e-mail to amsaa.reltools@us.army.mil.

1. System Reliability Model
2. Systems Engineering Integration
3. System-Level Operational and Environmental Life Cycle Loads
4. Life Cycle Loads on Subsystems, Assemblies, Subassemblies, and Components
5. Failure Modes and Mechanisms
6. Closed-Loop Failure-Mode Mitigation
7. Reliability Assessment
8. Reliability Verification
9. Failure Definitions
10. Technical Reviews
11. Methods and Tools
12. Outputs and Documentation

2.8.7. Modeling and Simulation (M&S)

One of the important M&S strategy decisions that must be made by the PM team early in a program is the allocation of M&S responsibility between the Government and its contractor(s), with attendant funding and accountability implications. This allocation typically varies by phase, with Government M&S activities prominent in the early phases (e.g., MSA and TD), and the prime contractor assuming a prominent role after source selection and throughout EMD. Government M&S activity typically increases again during OT&E.

The Government must decide to what degree it wishes to have an independent M&S-based capability rather than just insight into the contractor's M&S activities. The Government must also decide whether it will provide, or facilitate providing, the contractor with Government-owned M&S tools and data, and if so, what its limits of liability will be regarding the functional adequacy, trustworthiness, and evolution of such GFE or GFI. VV&A responsibilities must also be allocated. Close coordination is necessary between the program office's M&S lead and its KO.

Contracting strategies, solicitation, and contract clauses must be consistent with the decided division of responsibilities. Particular attention should be paid to the GFE/GFI aspects discussed above. RFP language and contract clauses should address M&S representation requirements; data rights; the contractor's own M&S planning and documentation, including the examination of reuse opportunities; expectations regarding the sources of M&S tools and data; the ownership and maintenance of Government-funded M&S resources; VV&A; standards that must be complied with; Government user support; access control; and metrics and documentation requirements, all across the system's full life cycle. A key planning consideration is addressing the need for including updates to M&S in the RFP based on use of actual test data. Effective use of M&S throughout the T&E process requires an iterative model-test-model process where possible.

Indicators of contractor M&S expertise should be considered in defining source selection criteria. Contractor attributes that have a direct relationship to successful M&S use may include the following:

- A documented systems engineering process showing its organizations, activities, the specific M&S tools used by each, and the information flows among them.
- An existing information-sharing infrastructure (e.g., integrated data environment) providing enterprise team members, on a nearly continuous, from-the-desktop basis, the capability to discover, access, understand, and download a comprehensive set of authoritative, accurate, and coherent product development information. The data items provided by this system should be accompanied by metadata providing the pedigree and sufficient applicability and context information to guide their valid use.
- Successful experience using a wide variety of M&S, both for design (prescriptive modeling environments such as systems engineering tools, computer-aided design, and software design tools) and assessment (descriptive M&S), from the engineering to mission levels.
- Successful participation in distributed simulation federations using an open standard architecture (e.g., the Institute of Electrical and Electronic Engineers Standard 1516 High Level Architecture).
- A record of reuse of M&S tools and information produced by other organizations (such as Government, industry, and COTS).

- A documented VV&A process, with records indicating a history of compliance.
- A staff with documented M&S expertise.

2.8.8. Government-Furnished Equipment (GFE)

The identification of and control for GFE for T&E must be identified early because both issues will affect contract funding and scheduling. In areas like support equipment, not identifying GFE can be a showstopper if an incorrect assumption is made about equipment availability. Similarly, the Government does not want to pay for development of contractor-unique support equipment if the design can use existing support equipment.

2.8.9. Safety

The type of product/system will drive the personal and system safety issues. Because the T&E program will involve real people using real systems, the strategy for ensuring the safe conduct of the test program must be captured. One issue of particular importance is where the final safety decision rests – with the Government (such as the program office or range safety officer) or contractor. Safety topics include accountability in case of an accident and weapon release authority. The solicitation should address how the contractor will provide technical data and drawings to Government safety offices to facilitate system safety evaluations and range clearances. The contractor must provide the Government with a safety assessment report and all associated material safety data sheets in accordance with MIL-STD 882, "DoD Standard Practices for System Safety" (Reference (z)).

2.8.10. Software

Software is a rapidly evolving, emerging technology that can now be found in major components and critical subsystems of most DoD materiel solutions. Software allows creation of products that fundamentally differ from hardware components. Differences between hardware and software include the following:

- Software has no physical characteristics limiting size or prescribing natural, structural units with boundaries and proximal interfaces.
- Software structural units are statements, objects, and programs for which the interfaces are intangible and range widely in diversity, complexity, and dynamic behavior.
- Software units are captured abstractions of functions allocated to design, easily changeable, and therefore challenging to manage and maintain.
- Unlike hardware that typically degrades gradually before failing, software typically fails abruptly and with greater consequence to delivery of expected system performance.
- Software almost always delivers function through code execution in a nondeterministic domain space and therefore cannot be exhaustively tested and will always contain faults. Software testing mitigates the risk of performance failures by exposing code faults and is therefore fundamentally a risk-reduction activity.

Software component implementations have the following distinct properties that make engineering and programmatic management inherently difficult throughout the system life cycle:

- **Complexity:** Difficulty in describing software structure and predicting behavior.
- **Changeability:** Having no physical properties; software can be easily changed throughout development and fielded service. Software change is inevitable, enabling responsiveness to changing threats, capability needs, technology advances, design improvements and corrections, and management resource budgets. Software changes may induce risks for which planning may be required.
- **Invisibility:** Without physical form, software architectural representations fall short of complete representation of complexity, size, and critical characteristics.
- **Conformity:** Software change is the means by which systems maintain conformance to changing service environments, management and resource constraints, and interfaces with hardware and other software systems. Conformity is achieved through near-continuous verification.

System designs that incorporate software components require consideration of these unique differences and their implications for software T&E. The requirement to demonstrate comprehensive software T&E capacity should be integrated into solicitations. Responses to the software T&E requirement should be evaluated in proposals, and past performance artifacts should be examined to address the following critical areas:

- Allocation of sufficient financial and schedule budgets, material, and domain expertise across the WBS and IMP/IMS to properly incorporate software T&E with software architecture and design development, software production, subsystem and system integration, and product sustainment.
- A comprehensive software T&E approach that specifically includes evaluation of high-risk technologies in system designs and complexity in the system software architecture. This approach should identify and describe:
  - Metrics and evaluation data for resource management, software system requirements, and product quality, including reliability and product reliability growth.
  - Types and methods of software testing to achieve comprehensive evaluation.
  - Software T&E directly supportive of the program risk management enterprise, and responsive to risk-reduction strategies and risk-mitigation activities.
  - Data management, analysis, and evaluation methods and tools.
  - Models and simulations contributing to software T&E, including accreditation status and planning. See DoD Directive 5000.59, "DoD Modeling and Simulation (M&S) Management" (Reference (aa).
  - Software development, integration and test, and software-hardware integration labs and facilities. See MIL-HNDBK-881, "Work Breakdown Structure for Defense Materiel Items Reference (bb).

- A defined software T&E process consistent with and complementing software unit, subsystem, and system development, maintenance, and systems engineering processes, committed to CPI and aligned to support project phases and reviews, including an organizational and information flow hierarchy.
- Software test planning and test design initiated in the early stages of functional baseline definition and iteratively refined with T&E execution throughout allocated baseline development, product baseline component construction and integration, system qualification, and in-service product sustainment. The solicitation should include resources, as appropriate, needed to complete DIACAP C&A according to Reference (q).
- Thorough T&E of design reuse (COTS, Government off-the-shelf) of software code, databases, and hardware, and associated test procedures or test data. Reuse planning should include a defined process for component assessment and selection, and T&E of component integration and functionality with newly constructed system elements.

3. SOLICITATION

The contents of this section will focus on and consider the most important contractual T&E items during transition from the pre-solicitation phase to the actual drafting of the RFP. In contracting, the term "solicitation" means to go out to prospective bidders and request their response to an RFP. The solicitation builds upon the SOO and the SOW. All previous identification, development, and refinement of T&E requirements now have to be captured clearly, completely, and accurately in the appropriate sections of the RFP.

3.1. Section B of the RFP: Supplies or Services and Process/Costs

Section B of the RFP contains a brief description of the supplies or services; for example, item number, national stock number/part number if applicable, nouns, nomenclature, and quantities, and includes incidental deliverables such as manuals and reports. All CDRLs should be reviewed for inclusion of T&E execution support (i.e., data rights, test data, test plans, source code drop, prototype quantity, delivery times/location).

3.2. Section C of the RFP: Description/SOW

Section C of the RFP contains the detailed description of the products to be delivered or the work to be performed under the contract. This section typically includes the Government's SOO (or SOW) and preliminary SPS. The preliminary SPS was addressed in section 3.13.1. Other documents may be included, such as sample IMP event descriptions, CDRL, Contract Security Classification Specification (DD Form 254), and pricing matrices. A major discussion item is the inclusion of the implementation and execution of reliability activities fully integrating systems engineering, DT, OT, IA, security, interoperability, and DIACAP C&A. Appendix F provides a checklist to guide your discussions and decisions relative to RAM planning, accountability, and reporting for the program. Questions for consideration: *Are all requirements clearly defined and stated in performance-based terms? Are performance-based characteristics directly tied to program objectives?*

3.2.1. Statement of Objectives (SOO)

The items in Table 2-2 should be addressed in a SOO to emphasize the main T&E themes of this guide. Specific program requirements and the program T&E approach will help tailor these items.

3.2.2. SOW/PWS

See section 2.7.2 of this guide for SOW/PWS content. The following five elements need to be considered during the proposal development:

- SOWs are often too detailed and inadvertently include inappropriate items for a contract. For example, technical day-to-day procedures and/or instructions are captured in such detail that as they mature during the program, they cannot be implemented without a contract change. The goal is to secure a commitment to implementing the process and not controlling the detailed procedures. The TEMP should capture how the T&E processes operate for the program. Therefore, the

SOW should refer to the TEMP as a commitment to implementing the processes defined for the program.

- SOW tasks should be reflected in the IMP/IMS, especially the technical baseline management, technical design, verification, and validation tasks and their associated system-level event-based technical reviews.
- The SOW should not identify individuals or specific IPTs that accomplish the tasks and should avoid including start dates or completion dates. These dates, and sometimes the IPTs that will accomplish the tasks, are identified in the IMS.
- Conducting event-based technical and test reviews should be appropriate and consistent with the program technical and support strategy included in the offeror's proposal.
- All the important T&E management processes and tasks should be included, such as decision analysis, T&E planning, assessment, test plans, reports, data requirements, and risk and configuration management. The information in the Appendixes can be a useful aid during the SOW evaluation to ensure completeness.

3.2.3. Test and Evaluation Master Plan (TEMP)

The TEMP is used to evaluate the completeness of program planning and application of T&E best practices. The following is a list of five considerations when evaluating the offeror's proposed integration of its T&E solution and program technical approach with the management approach, which should be included in a source selection evaluation guide or other appropriate document:

- The proposed T&E solution incorporates best practices and processes that are mature, stable, and will be applied to the program. Any tailoring or modifications to the standard processes (as reflected in corporate procedures) are appropriate to the program and should not increase cost, schedule, or technical risk. The offeror has made a corporate commitment and implemented plans for continuous process improvement.
- Major T&E reviews in support of the program's technical reviews (such as the TRR, preliminary design review (PDR), and CDR are clearly identified.
- A single T&E authority for the program has been identified. The T&E team's roles and responsibilities within the offeror's proposed organization have been clearly defined and assigned.
- The skill, security clearance, experience level, and corporate commitment of key proposed T&E personnel have been ascertained. Plans for transition and personnel assignments are in place for a smooth ramp-up of work tasks without risk of delays. Sufficient manpower resources have been identified and are available to support the program.
- Key T&E processes critical to program success have been integrated with program management, and engineering processes reflect the T&E approach in the TEMP.

3.3. Section E of the RFP: Inspection and Acceptance

Section E of the RFP includes inspection, acceptance, quality assurance, and reliability requirements. A question for consideration is: *Has the acquisition team developed a tailored quality assurance surveillance plan to monitor contractor performance?* This section should describe the organization and procedures to perform the R&M task.

3.4. Section F of the RFP: Deliveries of Performance

In Section F of the RFP, the KO will specify the requirements for time, place, and method of delivery of performance. Questions for consideration: *Has the required number (sample size) of test articles been identified? Has a delivery location and schedule been established?* If you think you may want the contractor-acquired property, have the KO state in the solicitation and resulting contract that title to the contractor-acquired property will revert to the Government at the end of the contract. This section will identify the PM's desire to have contractor support personnel available to repair or provide reach-back of contractor's product during DT&E effort. Identify contractor property needed as spares during the testing.

3.5. Section H of the RFP: Special Contract Requirements

In Section H of the RFP, the KO will include a statement of any special contract requirements that are not included in Section I, Contract Clauses, or in other sections of the uniform contract format. All contract clauses for data delivery, Government property, rent-free Government property, and personnel qualifications should be reviewed. This information may reside in Section H or I or both sections.

3.6. Section I of the RFP: Contract Clauses

In Section I of the RFP, the KO shall include the clauses required by law and any additional clauses expected to be included in any resulting contract, if these clauses are not required in any other section within the uniform contract format. An index may be inserted if this section's format is particularly complex. All contract clauses for data delivery, Government property, rent-free Government property, and personnel qualifications should be reviewed. This information may reside within Section H or I or both sections.

3.7. Section J of the RFP: List of Attachments

In Section J of the RFP, the KO shall list the title, date, and number of pages for each attached document, exhibit, and other attachment. Cross-references to material in other sections may be inserted as appropriate. This section should identify whether the TEMP is releasable to the contractor. If so, make sure the TEMP is provided to the contractor and listed in this section. Documents released to the contractor should be reviewed for security classification. Those documents non-releasable to the public should have a distribution list established so they can be viewed by the companies performing the work.

3.8. Section K of the RFP: Representations, Certifications, and Other Statements of Offerors or Respondents

Section K of the RFP includes those solicitation provisions that require representations, certifications, or the submission of other information by offerors. Requests for certain certifications that support the T&E strategy should be reviewed.

3.9. Section L of the RFP: Instructions, Conditions, and Notices to Offerors or Respondents

Section L of the RFP describes in detail the contents of each volume of the proposal. Inserted within this section of the solicitation are provisions and other information and instructions not required elsewhere to guide the offerors or respondents in preparing proposals or responses to RFIs. Prospective offerors or respondents may be instructed to submit proposals or information in a specific format or several parts to facilitate the evaluation. The instructions may specify further organization of proposal or response parts, such as administrative, management, technical, past performance, and certified cost of pricing data.

If the contractor will provide oversight for another contractor or direct work to another contractor, this section will describe what measures are planned or have been taken to reduce or eliminate potential organizational conflicts of interest. Section L of the RFP will:

- Describe the contractor test management structure for T&E, experience of T&E staff, the predicted staffing levels, and the application of T&E best practices.
- Define the responsibilities of the contractor and the Government during test planning (include contractor testing, DT, and integrated testing).
- Describe contractor's approach on technical data, including management, ownership, control, timely access, and delivery of T&E data, to include raw test data, to support the evolving technical baseline.
- Define CDRL and DIDs.
- Identify any T&E related data products that the contractor must provide.
- Determine applicability of DIDs in support of T&E efforts.
- Determine applicability of commercial certifications of material or product. *Does the RFP contain a top-level schedule depicting key T&E events?*
- Describe the allocation of M&S responsibilities between the Government and contractor and the expectations regarding the sources of M&S tools. *Has the acquisition team identified an industry day?*
- Define release of T&E assessment data to industry. *Is the program or aspects of said program classified? If so, is contractor capable of storing, handling, obtaining, and controlling classified data? Are contractor T&E personnel cleared to review the program?* Contractor should provide certification of classification capability along with designated personnel. *Is the acquisition team providing a copy of or access to the program TEMP or T&E strategy?*

3.10. Section M of the RFP: Evaluation Factors for Award

A successful offeror's proposal must respond to the requirements of the RFP. The proposal must be responsive to and consistent with Section L, Instructions, Conditions, and Notices to Offerors or Respondents. Section M is the standard against which the proposal will be evaluated and forms the basis for selection. To a large extent, the quality of the proposal is directly related to the clarity of the Government's delineation of the technical requirements in the RFP. During the proposal evaluation, the Government team will establish the degree to which the contractor has implemented RFP requirements and proposed a sound technical program with high expectations for success. Table 3-1 provides a summary of eight T&E focus and evaluation areas to consider in the Section M evaluation factors. This list is not meant to be all-inclusive; however, the acquisition team should limit the number of evaluation factors to focus attention on areas most likely to be discriminators among proposals. The inclusion of too many or overly detailed evaluation factors will consume source selection resources without benefit to the source selection process. DoD Components and programs may have specific proposal evaluation criteria that are tailored to the unique circumstances of the acquisition program.

Section M consists of the evaluation factors and how the contractor will be graded. Identify all significant factors and any significant subfactors that will be considered in awarding the contract and their relative importance. *Has the acquisition team mapped Sections L and M to the program supporting documents (AS/acquisition plan, TEMP and SSP) and requirements document (ICD, CDD, or CPD). Are minimum thresholds and maximum performance objectives clearly defined? Are requirements stated in certain terms such that evaluators will be able to assess whether the offeror meets or exceeds a particular outcome? What are the measures and metrics to evaluate qualification of contractor T&E personnel? Are critical program objectives reflected in the evaluation criteria?* Many of the documents in the RFP evolve into the negotiated contract via the proposal and source selection process (see Figure 3-1).

Table 3-1 T&E Focus Areas

- **T&E Best Practices**
  - Offeror addresses the T&E approach across the program life cycle.
  - Offeror has proposed event-based tests and reviews with entry, exit, and measure of success criteria.
  - Reviews include participation by Government and industry T&E SMEs.
- **Offeror's Capability**
  - Offeror's domain experience (process and product) is applicable to the program.
  - Domain expertise is combined with application of offeror's best practices using experienced personnel.
  - Offeror demonstrates proven, positive past performance (in domain and process areas) that supports a high probability of T&E success on the program.
  - Offeror provides an acceptable DR process and database compatible with the

Government's DR data requirements and database.

- **T&E Planning**
  - Adherence and application of corporate best T&E practices is inherent in the T&E approach.
  - T&E processes are integrated within the management and technical framework.
  - OT&E and JITC requirements are addressed (such as critical operational criteria, IA, SoS interfaces within the SoS and outside systems, and critical mission function).
  - Experimental design and analysis (i.e., DOE) and reliability growth planning processes and practices are used.
- **T&E Baseline**
  - Processes and resources (people, test ranges/facilities, instrumentation, and domain infrastructure) are integrated to systematically mature the T&E performance baseline without duplication.
  - Requirements management and traceability processes support the evolving T&E performance baseline
- **Metrics**
  - Product metrics are linked with T&E performance baseline maturity.
- **Award Incentives**
  - Award incentives support maturing the T&E baseline and are linked to quality and delivery performance of the final product.
- **Cost and Schedule Realism**
  - Program budgets and cost estimates are realistic. Cost, schedule, and performance are balanced.
  - Cost estimates and schedule support the T&E strategy and approach in the TEMP.
  - The program's critical path is actively managed.
- **T&E Data Access**
  - Ownership, control, timely access, and delivery of T&E data, including raw test data, to support the evolving technical baseline are clearly established. T&E data are consistent with the program's technical and acquisition strategies.
- **Cost-Effective Strategy Aligned With Integrated DT/OT Test Continuum**

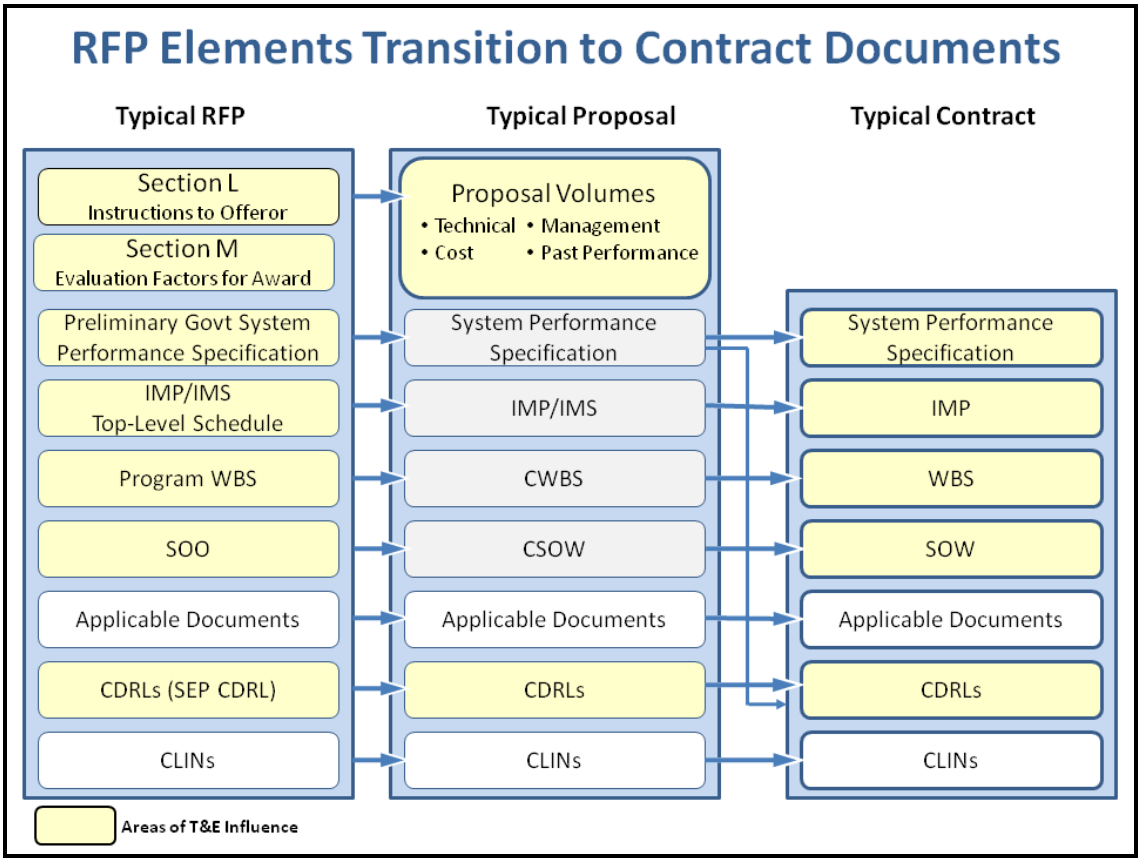


Figure 3-1 Relationship of Proposal Documents to Contract Documents

During the proposal evaluation, it is important that any changes or deficiencies in the proposal documents be corrected. The SSP delineates how the Government and the contractors will communicate during the evaluation process; for example, procedures for submitting questions or requests for clarifications and submitting a final proposal revision. The technical authority must ensure that any potential contractual documents are complete and sufficient. Usually the IMP, WBS, SPS, SOW, and CDRL are identified as contractual documents. Contract DIDs and CDRLs may be tailored to the acquisition program to obtain contractor-produced plans or studies that satisfy specific program needs. If the Government is expecting or relying on a contractor report to satisfy an acquisition milestone or decision review, then the CDRL should reflect a report delivery date in advance of the applicable review.

3.11. Overview of T&E Requirements

Sections L and M must capture the major thrusts of the T&E requirements described in the TEMP and other relevant T&E and acquisition documents. Section L of the RFP instructs the offerors on structuring their proposal and outlines what should be included in each section of the submittal. It should be written after Section M, and tracked to the evaluation factors. Sections L and M must mirror the SSP. The Government should avoid asking for unnecessary data in the proposal to satisfy technical curiosity because extraneous requests could cause the contractor's proposal team and the Government evaluation team to spend valuable time on areas not germane to the evaluation criteria.

The offerors will treat all data as critical. All data submitted in the proposal must correlate with the evaluation criteria in Section M or be necessary to award the contract (e.g., model contract, SOW, CDRL, SPS). If the offeror's time and resources are wasted on unnecessary data, the quality of the proposal may suffer, potentially affecting the choice of the right contractor with the right approach.

3.11.1. Integrated Master Plan (IMP)/Integrated Master Schedule (IMS)

The RFP should contain an event-based, top-level schedule depicting the major program elements and key milestones, such as contract award, test phases, reviews, production, long-lead decisions, and system delivery.

The IMP and IMS should clearly demonstrate that the program is structured to be executable within schedule and cost constraints, and with acceptable risk. They should provide a functionally integrated picture of the proposed program. There must be a direct correlation between the event-driven activities in the IMP and IMS and the planned technical approach. Thus, the IMP and IMS are key elements to proposal preparation and source selection. There must be a high correlation between the cost basis of estimates (BOEs) and information within the IMS. Table 3-2 provides a sample RFP Section L for the IMP/IMS.

Table 3-2 T&E Content for RFP Section L – IMP/IMS

Section L – IMP/IMS
The offeror shall submit an IMP/IMS guide that is structured as an event-based planning document. Engineering reviews such as the system requirements review (SRR), system functional review (SFR), PDR, and CDR are typical. T&E shall support each review, as required, with appropriate performance data. The IMP includes the accomplishments and criteria for the efforts involved with the design, development, test, production and sustainment including planned block upgrades, technology insertion, and entry and exit criteria. The offeror's T&E processes and corporate best practices (as described for the program) shall be the source of the test events, definitions, major T&E products, and criteria for the IMP events. The program's critical path is identified in the IMS. The result of a schedule risk assessment is presented and reflects acceptable schedule risk. For programs that require an IMP that includes a process narrative section (IMP-IMS Guide, section 4.2.5): The offeror shall include within the IMP process narratives brief synopses of the offeror's processes considered essential for program success. The narratives shall reference the offeror's corporate T&E processes and best practices and indicate how they are applied to the program.

3.11.2. Management Volume

The management volume is used to highlight special areas that are discriminators for source selection. It should not be used to systematically address all technical and management processes to be used on the program. It should, however, provide a clear description of how the offeror plans to organize internally, interface with the Government program office and other external organizations, and manage subcontractors. This volume should include the approach to managing all program information (including T&E), its assembly and integration, and its dissemination among stakeholders.

The proposal instructions should avoid a reliance on a "cookbook" list of specific T&E management processes to be discussed and evaluated. The important issue is that the offeror's T&E processes and best practices are mature, integrated, and will be applied to the program. The focus should be on the key T&E processes that are important for program success. Examples of discriminating processes for a program might include an experimental design and analysis approach (i.e., DOE analysis), HSI and usability analysis, DR, DIACAP C&A according to Reference (q), T&E KPPs, COIs, critical operational criteria, CTPs, metrics and system reliability growth, software maturation, program and performance review processes, and M&S. Table 3-3 provides a sample Section L for the Management Volume.

Table 3-3 T&E Contents for Section L – Management Volume

Section L – Management Volume
The offeror shall submit a Management Volume that describes the key management and technical processes and their integration with the other management, financial, and functional processes. This volume shall include discussion of processes, program organization, and special tools that are important to technical management; for example, program organization, and roles and responsibilities of IPTs and the T&E team. The volume shall include T&E requirements management tracking tools, electronic and/or virtual program approach, special capabilities/facilities, data management/archiving/real-time access and data submittal, configuration management and supporting tools, M&S processes, and risk management processes. The volume shall include the role of reviews in baseline management, and system validation and verification processes including failure/fix reporting and tracking.

3.11.3. Contract Data Requirements List (CDRL) and Data Item Descriptions (DIDs)

CDRLs and DIDs may be tailored to the acquisition program to obtain the following contractor-produced documents that satisfy specific program needs:

- **CDRL**. In this section, any T&E-related data products that the potential contractor must produce are identified. This may include plans, metrics, reports, artifacts, raw test data, or other T&E documentation. The CDRL will delineate the specific M&S items and data products, and the timelines to provide these to the designated test organizations.
- **DID**. Any DIDs applicable to the T&E effort should be included in this section. A DID is a completed document that defines the data required of a contractor. The document specifically defines the data content, format, and intended use. Each T&E team will have to determine the need for DIDs supporting their effort. To determine whether a T&E DID already exists, go to the Acquisition Streamlining and Standardization Information System (ASSIST) Website at https://assist.daps.dla.mil/online/start/. ASSIST is the source of DoD specifications and standards. Examples of T&E DIDs are:
  - **DI-NDTI-80566A – Test Plan**. The test plan outlines the plans and performance objectives at every level of testing on systems or equipment. It provides the procuring activity with the test concept, objectives and requirements to be satisfied, test methods, elements, responsible activities associated with the testing, and the required measures and recording procedures to be used.
  - **DI-NDTI-80809B – Test/Inspection Report**. This DID contains the format and content preparation instructions for the data product generated by the specific and discrete task requirement as delineated in the contract.
  - **DI-NDTI-81585A – Reliability Test Plan**. This plan describes the overall reliability test planning and its total integrated test requirements. It outlines required reliability tests, their purpose, and schedule. This document will be used by the procuring activity for review, approval, and subsequent surveillance and evaluation of the contractor's reliability test program.

3.12. General Factors

To accommodate variations among the DoD Components' source selection processes, RFP format nuances, and differences among programs, the discussion of Sections M and L is segmented into four general factors: (1) management, (2) cost factor, (3) past performance factor, and (4) cost factor or pricing data. Each of these areas includes a brief discussion of the topic and sample language (in shaded boxes) that can be applied to program RFPs.

Section M of the RFP states the evaluation factors and significant subfactors (and their relative importance) that are the basis for selecting the source. Section M should be written before Section L, and should be carefully structured to address only those elements determined to be keys to success. Taking into account early industry input, focus the Section M criteria on the source selection discriminators required to select the best value proposal with acceptable program risk. Do not include proposal evaluation criteria that do not add value to the source selection. Weigh each lesson learned from previous programs and RFPs (especially similar programs) when establishing RFP requirements.

Sections M and L should be specific to each program, giving consideration to the scope and the nature of the technical program, maturity of the relevant technology, critical subcontract or teaming efforts, software content, and COTS/non-developmental item. The task for the Government team is to provide the one-for-one match between the Section M criteria that will be used to evaluate the technical information and the proposal instructions in Section L. Normally, there are three primary considerations:

- Offeror's plans for implementing and managing the T&E process.
- Offeror's technical approaches (program and specific product offering), including supporting data (trades and analyses).
- Offeror's past performance.

The most effective criteria are measurable and relevant to the program, traceable, and under the offeror's control. The Government team should answer the following questions when developing specific program-related criteria for Sections M and L:

- How can the evaluation team develop confidence that the offeror's proposed T&E solutions, including unprecedented high-risk solutions (e.g., lack of proven technical maturity), will effectively measure performance and can be implemented within technology, cost, and schedule baselines?
- How will the evaluation team establish an understanding of the offeror's T&E approach?
- How can the evaluation team understand whether the specific plans for implementing and managing the T&E processes were based on company best practices, domain experience, and company maturity ratings?
- How will the evaluation team understand whether the T&E solution is adequately supported by trade studies, LDTs, analyses, M&S, and demonstrations? How will the evaluation team determine whether the supporting trade studies, LDTs, trade criteria, and analyses are the results of the T&E process during proposal preparation? Is there objective evidence that the offeror used the processes proposed for the program?
- How will the evaluation team determine that relevant and demonstrated past performance from other programs is applicable to the T&E processes for the proposed approach (e.g., successful performance on similar complex systems)?
- How will the evaluation team assess the maturity and application of the offeror's proposed processes in the proposal risk assessment?
- How will the evaluation team determine that the T&E costs and resources (especially number of operators, sample size, tests, ranges, and usage schedule and sequence) proposed for the system/subsystems are reasonable and realistic for the planned T&E approach? Is a scientifically based test design process such as DOE used?
- How will the evaluation team establish that the offeror's proposed T&E schedule and critical path analysis are realistic and represent the planned T&E approach consistent with the overall program schedule?

- How can the evaluation team understand the trustworthiness of any M&S proposed for use in the T&E process?

It is common practice to include a matrix in the RFP that correlates Section L to Section M so it is perfectly clear what portions of the proposal will be used to evaluate each Section M evaluation criteria element. Doing so also serves as a quick check to make sure that each element of the proposal tracks to source selection criteria. Figures 3-4 through 3-8 include sample content for Sections M and L for each subject that needs to be integrated with the rest of the information in Sections M and L in the program's RFP.

3.13. Technical Factors

T&E team members should be involved in the review and assessment of the technical portions of the source selection. This review generally involves the following:

- The offeror's proposed technical solution.
- The technical data supporting the offeror's proposed technical solution and how it meets the specification requirements.
- The SPS (or equivalent).
- The review document, the TEMP, or equivalent.

The core of the technical evaluation centers on the offeror's SPS, the technical solution of the approach, and any supporting trade studies, LDTs, analyses, modeling, and demonstrations that have been requested in Section L.

Most RFPs request two general types of technical data: the description of the proposed solution, and trade studies and analyses. The proposed solution and resulting performance are program specific and represent the bulk of the technical data submitted. This section includes drawings, flow diagrams, technical descriptions, and illustrations or photographs of the offeror's proposed solution. This important information is, in essence, the result of the engineering processes to include DT&E processes implemented by the bidder during the proposal phase.

The trade studies and analyses (including M&S) provide substantiating data showing not only the performance but also the extent and scope of alternative solutions considered before arriving at the proposed solution and specification. A well-structured family of trade studies, analyses, and M&S that supports the system configuration and its performance is objective evidence that the bidder has implemented its engineering processes described in other sections of the proposal. The Government should ask for a summary of the trade studies, LDTs, and analyses that discuss the scope of the alternative solutions and performance capability considered before arriving at the proposed solution and specification. Many times "why" something was discarded is as important as "what" was selected. The trade study, LDTs, and analysis data clarify the inner workings of the offeror's processes. The data demonstrate the application of the offeror's requirements analysis process and are evidence that the offeror:

- Has engineering and T&E processes.

- Has applied them in arriving at a solution.
- When coupled with other documents in the proposal, is committed to continue the processes during execution of the contract.

Tables 3-4 and 3-5 provide sample content for Sections M and L for the supporting T&E data, which need to be integrated with the program-unique parts of Sections M and L.

Table 3-4 T&E Content for Section M – Supporting T&E Data

Section M – Supporting T&E Data
This supporting T&E data factor (subfactor) is met when the offeror's proposal demonstrates the following: 1. The offeror conducted a series of trade studies, LDTs, M&S, and analyses that systematically evaluated the full range of alternatives. The results support the technical and program requirements and validate the proposed configuration and its performance. 2. Trade study and LDT processes were uniformly and consistently applied and followed the offeror's documented corporate processes as applied to the program in the TEMP. 3. Trade study and LDT criteria addressed the critical cost, technology, risk, and performance requirements/constraints for the program. 4. Recognition that an RPP is required to understand Government requirements and the need to design and test for product/system reliability.

Table 3-5 T&E Content for Section L – Supporting T&E Data

Section L – Supporting T&E Data
The offeror shall provide a summary of the T&E trade studies, LDTs, M&S results, and product/system reliability and analyses that were accomplished to arrive at the proposed solution. The offeror shall discuss the approach to the following topics: 1. The trade studies, LDTs, analyses, and M&S processes. 2. A summary of the trade studies and LDT results that support the proposed solution and program T&E approach. 3. A description of the trade study and LDT criteria, their relation to the key performance requirements/constraints for the program, and the planned processes addressed in the TEMP. The data shall address the range of alternatives considered and the important results that support the T&E strategy and approach decisions. 4. The process for developing and implementing an RPP.

3.13.1. System Performance Specification (SPS)

A preliminary SPS that defines the Government's performance requirements for the system is normally included in the RFP. The offeror normally responds with an SPS in the proposal. This specification includes the Government requirements plus any derived requirements necessary to describe the system-level performance. It may include allocation of requirements and should include corresponding verification requirements. The SPS should not include SOW language, tasks, guidance, and data requirements but should reference necessary industry and approved military specifications and standards.

Offerors responding to the RFP have a tendency to parrot back the Government's preliminary SPS in the proposal. They are hesitant to revise the content and format and are especially hesitant to respond with revised requirements for fear of being judged nonresponsive. The Government should make clear in the solicitation that the offerors need to do so. The RFP should clearly delineate whether the Government is receptive to considering revised, cost-effective performance requirements (trade space), along with an indication of how the value to the Government will be established and evaluated. The system specification will be included in the contract.

In past practice, one particular element of the system specification has received limited emphasis—section 4.0, Verification and Test. The offeror must supply more than a simple table indicating the method of verification (analysis, inspection, simulation, test, or demonstration). Section 4.0 of the specification, along with the system test plan, IMP/IMS and TES/TEMP, should provide the insight to understand the method and extent of system verification. An incremental buildup approach to testing, including the T&E success criteria for each increment starting at subsystems of the system hierarchy, should support minimizing the system test events and activities. Section 4.0 of the system specification should reflect this T&E philosophy. Tables 3-6 and 3-7 provide sample content for Sections M and L for the SPS. These samples should be modified for the program and integrated with the rest of the RFP's Section M.

Table 3-6 T&E Content for RFP Section M – SPS

Section M – SPS
The offeror's SPS will be evaluated in conjunction with the technical solution based upon the following: 1. Specification includes the key requirements and functionality identified in the RFP's preliminary SPS stated in performance terms. 2. Requirements are verifiable by test, analysis, demonstration, or inspection and are supported by sufficiently mature technology (as defined by a technology readiness level (TRL) for the type of system under development, and dependent on the type and phase of acquisition (i.e., for an advanced concept technology demonstration, pre-MS B, exploratory technology). 3. Objective values (goals) are clearly identified and distinguished from firm threshold requirements.

4. Operational environment is described and defined in which the system, SoS, and/or family of systems (FoS) operates.
5. Environmental and safety design requirements and/or constraints are specified.
6. Functional and physical interfaces for the system are included.
7. Government and industry specifications, standards, and guides are used appropriately. Only approved Government documents should be referenced.
8. Test, verification, and reliability approaches for all system requirements included in the specification are complete and appropriate.
9. The specification does not include unnecessary requirements/language. (Examples include SOW tasks, data requirements, product or solution descriptions.)
10. The requirements are achievable within the planned program schedule and cost.

Table 3-7 T&E Content for Section L – SPS

Section L – SPS

The offeror shall propose an SPS that meets the Government minimum requirements. The specification should be performance based and address the allocation of Government performance requirements plus any derived requirements necessary to describe the performance of the integrated system solution. It should not be a mere "parroting back" of the Government's preliminary SPS but keyed and tailored to the individual solution of the offeror. Key elements to be addressed in the SPS are as follows:

1. Accurate and complete understanding of the key performance requirements (e.g., KPPs) in the Government's preliminary SPS included in the RFP.
2. Derived requirements necessary to document the system performance that will govern the design, development, and test program (e.g., CTPs).
3. Identified and documented system-level interfaces that define the operational, physical, and functional interfaces that define the program external interfaces and constraints (e.g., approved operational, functional, and/or system architectures).
4. Verification section to the specification that delineates the approach to verifying performance, success criteria, and key characteristics, including reliability metrics.
5. A cross-reference matrix showing the tracking of Government performance requirements to the offeror's proposed SPS. The specification should be structured for the proposed system solution and not restricted by the structure of the Government's preliminary SPS. In general, the offerors follow the structure and organization of the Government preliminary SPS when preparing the proposal's SPS. This may lead to an awkward specification structure if the offeror's breakout of the product differs from the Government's top-level breakout. It should be clear in Section L that the format of the Government preliminary SPS is to be followed or that the offeror has the latitude to restructure the specification to conform to its proposed technical approach.

As discussed in section 2 of this guide, the source selection technical evaluation is primarily focused on the offeror's proposed SPS, product-offering technical solution description, and supporting data.

The following 11 areas need to be considered during the technical performance proposal evaluation and must be consistent with evaluation criteria contained in Section M:

- All the critical or key requirements must be included within the specification.
- Goals are appropriately identified and differentiated from firm requirements. Goals do not have as much standing as contract performance requirements.
- Specification requirements are stated in performance language.
- SOW tasks or data deliveries are not in the specification.
- The SPS verification and test section (section 4) should be more detailed than a table reflecting only a method of verification. There should be a one-to-one correlation with the performance requirements (section 3), and it must reflect the engineering and test approach documented in other sections of the proposal.
- System hardware and software interface requirements should be identified and documented. They become critically important constraints on the system.
- Watch for "parroting" of the Government requirements without regard to substantiating evidence in the other sections of the proposal. A claim of performance without substantiating data is a technical risk.
- The product offering is complete, meets performance requirements, and is supported by hardware and software demonstrated in a relevant operational environment.
- The product reflects special design considerations such as MOSA, safety, and security.
- Analyses, M&S, and trade studies support design decisions and technical approach to the program as defined in the offeror's T&E approach.
- The processes should systematically address the technical challenge. The effort should be comprehensive (e.g., include all relevant solutions, technologies, and/or alternatives) and address the areas of technical performance, cost, schedule, and risk.

3.13.2. Management Factor

T&E management, design, integration, and verification/validation processes are normally evaluated using a combination of the offeror's SOW, TEMP, IMP/IMS, and management volume, as directed to be submitted with the proposal. The purpose of the evaluation is to establish the following:

- The offeror's domain current and past performance and experience.
- The stability and maturity of the offeror's T&E processes and best practices.

- That valid and complete approaches to test and evaluate the proposed system/subsystem are consistently integrated throughout the program

Table 3-8 provides sample content for Section M for technical and management integration. An integrated example is provided because there is significant overlap of all these elements. Individual Section L examples are included within each subsection.

Table 3-8 T&E Content for Section M – Technical and Management Integration

Section M – Technical and Management Integration

This factor (subfactor) is met when the offeror's proposal demonstrates the following:

1. The program tasks are complete and include a comprehensive description of the engineering and test tasks. Technical and test planning is complete, supports implementation of the program's technical strategy, and supports accomplishment of the requirements and objectives contained in the proposed contract. Plan for the management of technical and performance baselines and requirements using a tool set applicable to the program.
2. T&E processes are mature, stable, and represent the program's application of corporate best practices and lessons learned.
3. Approach, tasks, processes, and procedures are flowed down to the subcontractors, vendors, and other teammates. A trained workforce (familiar with the processes, practices, procedures, and tools) is available or in place to ensure accomplishment of the work.
4. T&E processes, products, and events are included in the IMP/IMS and reflect the program technical approach. The IMP narratives include the T&E processes and subprocesses; for example, requirements management and tracking, performance baseline control, interface management, configuration management, test data management, validation and verification process, failure reporting and corrective action system, and risk management.
5. The IMS clearly indicates the program's critical path and has acceptable schedule risk.
6. The T&E meetings, test events, status reviews, and design reviews are identified, participation is established, and schedule is set up to monitor and control T&E progress and support the technical progress.
7. A single T&E authority is responsible for program T&E direction with lines of responsibility and authority clearly established. Key personnel are assigned and personnel resources are identified. The role of the Government (program office, supporting Government organizations, and user) along with the key subcontractors has been identified.
8. Computer-based or software tools that are used for T&E management are real time (or near real time) and accessible to all program participants. Processes, procedures, and tools for test data archiving and data deliveries are secure and accessible to appropriate program participants. The tasks, activities, and methods are in place to facilitate the Government's timely access to the necessary program T&E.

9. System-level T&E reviews and meetings are adequate to monitor and control T&E progress in support of the technical progress. IMP events include T&E milestones consistent with the technical and support strategy for the program. The approach to event-based reviews is sound.
10. Evaluation product metrics address the key product performance requirements. The "leading and lagging" metrics provide past progress, current status to aid day-to-day management of the program for timely decision, and future projections. Root cause analysis processes are in place to continually improve the T&E processes and subprocesses. Tracking and reporting T&E progress and performance metrics at major program reviews are in place to ensure consistent application and continuing maturity of essential program processes (technical, configuration and data management, quality, subcontractor management, manufacturing, risk management, test and verification).
11. Program working groups are established that effectively involve program participants to improve coordination with supporting organizations and streamline T&E and other decision making.
12. The offeror's approach is based on corporate procedures and addresses the critical T&E areas within the program. The plans are flowed down to the teammates, subcontractors, and vendors involved in the program. The plans are consistent with the SOW, SEP, IMP/IMS, and other program management plans and processes to support critical path analysis, EVM, and risk management.
13. Along with strong technical, logistical, and contracting leadership, the program team has experienced T&E subject matter expertise as the program moves through its steps in contract formulation and execution. The Lead for T&E is involved with the KO in the program acquisition planning process as early as possible.

The management factor is typically evaluated using a combination of the offeror's SOW, IMP/IMS plus IMP narratives, and management volume.

3.13.3. Price or Cost Factor

Government source selection teams have placed more emphasis on evaluating the reasonableness of the offeror's proposed price or cost. Considerable emphasis has been placed on cost estimating, parametric analysis, BOEs, and the use of historical and past performance data on topics such as software code, hardware design complexity, T&E, and manufacturing costs. However, T&E tasks and costs have not been subject to the same analytical attention or scrutiny over the years. T&E personnel should consider the following five areas in support of the cost proposal evaluation:

- The T&E cost estimates correlate with the proposed solution and T&E program. Make sure that the program proposed is the one in the cost estimate and that it is reasonable and realistic. The program cost, schedule, and performance must be balanced and synchronized.
- The processes, organization, T&E tasks, and products proposed in other sections of the proposal are adequately resourced and included in the cost.

- Cost estimates for T&E work and products are supported by the offeror's domain experience and past performance.
- T&E manpower estimates and BOE must be adequate and reasonable for the organization, tasks, and schedule as reflected in the IMP/IMS and SOW. The skill level of the proposed manpower should reflect the complexity of the tasks. BOE supporting rationale should be based upon credible historical data, past experience, and/or expert judgment.
- Time phasing of the resources (manpower, facilities, and infrastructure) must be consistent with the IMP events and the IMS tasks and the TEMP's T&E approach.

Because costs are normally provided by WBS element, the program work breakdown structure (PWBS) is a valuable tool in understanding the cost proposal. The Government normally includes a PWBS based on Reference (bb) in the RFP. This PWBS must contain elements for T&E tasks along with the other elements (e.g., product, engineering, and sustainment). The RFP directs offerors to expand this Government PWBS into a contract WBS.

3.13.4. Past Performance

In a competitive environment, the Government relies upon the offeror's past performance record to demonstrate that the team possesses the skill and experience to perform well on the new contract. To gain this confidence, source selection groups, such as the Air Force Performance Confidence Assessment Group, use a structured approach driven by the respective source selection evaluation criteria to ensure that they fully understand each offeror's strengths, weaknesses, and risks. This, in turn, will allow the source selection team to project how those strengths and weaknesses will affect the proposed effort. T&E planning, leadership, and execution must have a prominent role in the Section M factors and subfactors and must be considered in the past performance evaluation. A contractor with experienced personnel in the applicable domain, bolstered with a credible past performance record, should result in better contract performance (e.g., lower risk and cost while still achieving the user's performance requirements). Table 3-9 shows a sample Section M, Table 3-10 provides an example of a rating scale, and Table 3-11 shows a sample Section L.

Table 3-9 T&E Concerns for Section M – Past Performance

Section M – Past Performance
The source selection group conducts a past performance assessment that evaluates the offeror's recent and relevant experience as a prime contractor or subcontractor, as well as the performance demonstrated by divisions and subcontractors that will participate in contract performance if the offeror's proposal is selected. Based on the assessment, the source selection group determines a confidence rating indicating the probable level of successful performance in planned effort, and identifies issues that may be a concern for the procurement.

Table 3-10 provides an example of typical past performance confidence assessment criteria and rating scale. DoD Components may have their own and more expansive assessment criteria, especially when considering C4ISR systems, SoS, or FoS experiences.

Table 3-10 Example of a Rating Scale for Past Performance

Performance Assessment Criteria	
Rating	**Description**
High Confidence	Based on the offeror's performance record, the Government has high confidence the offeror will successfully perform the required effort.
Significant Confidence	Based on the offeror's performance record, the Government has significant confidence the offeror will successfully perform the required effort.
Satisfactory Confidence	Based on the offeror's performance record, the Government has confidence the offeror will successfully perform the required effort. Normal contractor emphasis should preclude any problems.
Unknown Confidence	No performance record is identifiable.
Little Confidence	Based on the offeror's performance record, substantial doubt exists that the offeror will successfully perform the required effort.
No Confidence	Based on the offeror's performance record, extreme doubt exists that the offeror will successfully perform the required effort.

Table 3-11 T&E Concerns in Section L – Past Performance

Section L – Past Performance
A source selection group is convened to accomplish a performance risk assessment of offeror's relevant contract performance. The offeror's T&E performance record determines what level of confidence the source selection group has in the ability of each offeror to perform all aspects of the contract, to include T&E. Offerors must submit information on contracts considered relevant in demonstrating the ability to perform the proposed effort including rationale supporting the assertion of relevance. Section M evaluation factors and subfactors will be used to evaluate past performance and assess performance risk.

Most past performance assessments include a questionnaire that requests specific information relative to a contractor's past performance from selected previous customers of the offeror. Questions specifically for technical planning, leadership, T&E, and execution should be included when appropriate. See Appendix C for a sample past performance questionnaire.

Not all contracts included in the offeror's past performance volume need to be "highly relevant" to past performance, but a few examples should be highly relevant to the planned effort. See Subpart 15.305(a)(2) of Reference (e) regarding evaluating past performance and mandatory and discretionary requirements. Having limited T&E of a similar system, limited past performance results, or lack of domain experience can be a serious risk.

The T&E team needs to consider the following six areas in support of the past performance proposal evaluation:

- Focus on those contracts that are relevant or highly relevant and closely evaluate whether the performance is clearly applicable to the proposed program. Contracts that are similar in scope, apply the same corporate processes, and present successful results are the most powerful evidence of past performance.
- Review the allocation of T&E tasks to teammates and subcontractors and determine that their T&E experience is relevant and connected to the past performance examples.
- Most past performance evaluations include a questionnaire sent to select previous customers. Evaluate responses against the technical and management evaluation criteria in Section M.
- Systems engineering, and associated T&E, is a required element in Government acceptable contractor performance assessment reports (CPARs). This information is available to the past performance evaluation team. Trends and systemic issues across several contractor performance evaluations may indicate potential strengths and/or weaknesses in expected performance.
- For any past program evaluation rated as low, determine whether there is a "corrective action" plan between the Government and contractor and whether the corrective action is on schedule. Low contractor performance assessment rating with no corrective action plan is an indicator of risk.
- The team should evaluate not only the information provided by the offerors but information obtained from other sources (e.g., CPARs, questionnaires, internal Government information).

There are two Section J attachments for past performance: past performance questionnaire and previous contracting efforts. Previous contracting efforts are previous or ongoing acquisitions that are recent and relevant to the current acquisition. The past performance evaluation panel should access the Past Performance Information Retrieval System (PPIRS) at www.ppirs.gov for CPARs.

3.13.5. Proposal Risk Assessment: T&E Risks

Normally, the source selection team establishes a proposal risk for each of the factors established in Section M. The proposal risk is typically established at the factor level; for example, technical and management; however, the risks are identified at the subfactor level and summed to the factor during the evaluation. This risk assessment establishes the risk associated with the offeror's proposed program to include the technical approach, technical performance, testability and measurability of the performance requirements, management approach, application and integration of management and technical processes, program schedule, and cost/resource allocations. The following is a list of nine considerations when assessing the risks during the proposal risk assessment:

- Claims of performance are supported by credible analyses, trade studies, LDTs, and/or M&S results.
- The offeror's domain experience supports the program approach and the T&E challenges on the program.
- The T&E processes and best practices are mature and stable, and modifications to the standard processes (as reflected in corporate procedures) are appropriate to the program and should not increase cost, schedule, or technical risk.
- T&E processes are stable and mature, including technical hardware and software readiness levels (TRLs), maturity ratings (e.g., for MS B, a TRL of 6 is required), common discrepancy reporting criteria, and corporate plans for continued process improvement are in place.
- The key T&E processes determined critical to program success have been integrated into the program management and T&E approach. Examples include configuration management, requirements management, performance baseline control, risk management, technology insertion/obsolescence planning, data management planning, and M&S planning. These are flowed down to teammates, subcontractors, and vendors.
- The T&E processes, as appropriate, are integrated with the other functional processes (e.g., systems engineering, acquisition, and M&S).
- The risks associated with executing the T&E activities have been evaluated with respect to their relationship to the program's critical path.
- The risks associated with the offeror's costs are consistent with their proposed T&E effort, tasks and products, organization and personnel resources, and personnel experience levels.
- The T&E program schedule is reasonable and realistic and is consistent with the planned execution of the program; the T&E activities are on or near the program's critical path and are supported by the offeror's past performance.

4. CONTRACT EXECUTION

The contents of this section will focus on and consider the most important contractual T&E items transition from the solicitation phase to contract execution.

The keys to contract success are sound leadership, sound planning, and application of the contractor's corporate processes during execution. The T&E processes will develop, capture, document, and archive all of the T&E data. The T&E processes must be tightly integrated with the engineering and management processes and schedules that control the conduct of the program that will ultimately define, produce, and deliver the product to the user.

Program start-up can be hectic. New personnel are assigned, facilities are being established, and during all this turmoil, real program work needs to be accomplished. Program start-up and personnel ramp-up are almost always risk areas. It is essential that the program quickly transition into execution. During the first few weeks after contract award, it is important that the Government and contractor T&E team have an interactive face-to-face meeting, usually the kick-off meeting, and that the T&E leaders step forward and set the tone for the program. Focus areas during initial meetings with the contractor should include the following eight topics:

- Leadership completing the merger of the Government and contractor T&E personnel into a functioning integrated team; recognition of the responsibilities inherently residing with the contractor and Government (program office, user, evaluator, tester, and DCMA). T&E SMEs can participate in a variety of different teaming arrangements, including oversight teams, requirements teams, program management teams, and program-specific teams such as a CTF, CTT, or ITT. Regardless of the team's title, the team can have a T&E-specific focus, or not. The charter is the key document to define the team structure and should list the roles, responsibilities, products, and membership.
- Review of the program T&E strategy and approach and contractor and Government testing responsibilities.
- Review of the SPS, KPPs, and CTPs to ensure a mutual understanding of the functional baseline.
- Reinforcement of the importance of implementing the contractor's T&E best practices and domain experience.
- Review and establishment of the initial set of T&E product and process metrics.
- Review of the plans for event-based reviews (along with entry, exit, and measure of success criteria) documented in the IMP; review of the technical tasks and resulting products documented in the IMS; and ensuring T&E correlation with the SEP, IMP/IMS, and the EVMS in preparation for the integrated baseline review (IBR).
- Review and discussion of all the source selection T&E-related findings to ensure that they are resolved.

- Conduct of an IBR on contracts requiring compliance with DoD EVMS criteria requirements usually within 6 months after contract award.

4.1. T&E Team

At contract award, the Government/contractor T&E team must begin the integration into an organizational structure to promote the execution of the program's T&E processes and products. The authority for the Government and contractor must be clearly established. The contractor has likely identified a planned organizational structure in its proposal. The roles and responsibilities of Government personnel within the program's structure have to be defined and working relationships established. One of the first tasks is to make the appropriate assignments of Government personnel and to get the team physically together so introductions and working relationships can be established.

If the program organization includes a T&E WIPT, that team is often responsible for delivery of the completed TES or TEMP, JCIDS documents and operating concepts, and other T&E-related documents, and is responsible for the functioning of the T&E processes across the program. The team must be strong and staffed with experienced personnel from the Government and the contractor. The respective team uses the approved performance baseline (e.g., APB criteria) that is allocated to the product/system. The team is responsible for supporting the many major system reviews (SRR, SFR, PDR, CDR, etc.) with T&E results, and for risk assessments that will support the evolving technical baseline and product/system definition. Government participation on the respective teams is generally governed by the following eight guidelines:

- The Government does not lead or manage the contractor's T&E effort.
- Government participants serve primarily as "customer representatives," and one of their contributions is to reduce the cycle time of contractor/Government communications and decisions. The Government participants facilitate the Government's acquisition-related guidance and direction to meet program commitments in a timely manner.
- Government participants convey their knowledge/expertise on T&E approach, performance requirements, operations, maintenance, and other important topics.
- Government participants interface and coordinate the activities with other Government organizations that participate in the program, ensuring that they understand the overall T&E approach and that their participation supports program objectives.
- Government participants control and facilitate identification and delivery of GFE and Government-supplied data and services.
- Government participants should be participants in the risk management process.
- Government WIPT participants can offer personal and expert opinion from the customer's perspective; however, they cannot authorize any changes, waivers, or deviations to or from the contract requirements, which must be made by the KO.

- Government WIPT members cannot authorize contractors to perform work that is beyond the contract. Any such changes must be made by the KO.

4.2. Contractor Performance Information

Subpart 42.15 of Reference (e) identifies the requirement to record and maintain contractor performance information and requires the periodic assessment of contractor past performance. Up until 2002, most DoD Components used the CPAR. In July 2002, DoD endorsed the PPIRS as the single, authorized application to retrieve contractor performance information. PPIRS is a Web-enabled, enterprise application that provides timely and pertinent contractor past performance information to the DoD and Federal acquisition community for use in making source selection decisions. DoD Components should have some form of accepted documentation to record and maintain contractor performance information. Poor performance will influence source selection decisions and can result in non-selection. Excellent performance can significantly enhance the likelihood of winning a future source selection. Contractors are very sensitive to these facts and usually are motivated to improve poor performance. Used correctly and actively, contractor performance information can be an excellent management incentive tool.

4.3. Award Fee Implementation

There are several award fee activities that may require T&E involvement to sustain contractor and Government attention and interest in successful execution of the T&E approach to the program. These include interim and final evaluations for each award fee period, establishment of criteria for the upcoming terms, and providing feedback to Government officials and the contractor. It is particularly important to develop well-defined criteria applicable to each term, especially when an award fee is rolled over. (In rare cases, the fee determining official may agree to "roll over" unearned award fee money from one period to another to a subsequent term in accordance with Reference (u). The DFARS Service Supplements and Guides provide details regarding administration of award fee programs.

4.4. DCMA Support

The fundamental responsibilities of DCMA include the following:

- Assess compliance with contractual terms for cost, schedule, and technical performance in the areas of design, development, and production.
- Evaluate the adequacy and perform surveillance of contractor engineering efforts and management systems that relate to design, development, production, engineering changes, subcontractors, tests, management of engineering resources, reliability and maintainability, data control systems, configuration management, and independent research and development.

Because DCMA is normally on-site with the contractor, it is uniquely situated to be involved in the day-to-day contractor activities. DCMA is intimately familiar with the inner workings of the contractor's capability, processes, personnel, and facilities. It can be the "eyes and ears" of the program office and can be a valuable asset to the Government Lead for T&E. The KO or PM may negotiate an MOA with the DCMA field office detailing the specific tasks related to program participation after the contract is issued. (Many contract administration functions are routinely delegated to DCMA. See Subpart 42.302 of Reference (e) and Subpart 242.302 of Reference (d) for details. This activity should include how DCMA will participate in the execution of the T&E processes, and enlisting DCMA support in the implementation of various management tools/systems (WBS, IMP, IMS, EVM). The following three topics should be clearly addressed early in the T&E approach development effort, as appropriate, to the product/system under development:

- **Production Acceptance T&E**. DCMA usually is responsible for production acceptance testing. This responsibility and process should be verified and captured in the T&E process and approach.
- **Flight Release**. DCMA usually issues the flight release (in the case of aircraft programs) that permits even DT aircraft to enter the flight test program. This responsibility and process needs to be captured early in the T&E effort and schedule for the decision points that lead up to issuance of the flight release.
- **Contractor Personnel Management**. DCMA will sometimes be the approving authority for contractor flight crews to fly in DTs. This issue and the related DCMA processes and policies regarding training and certifying contractors to operate the system being developed must be captured early in the T&E process and approach.

For DCMA-specific responsibilities associated with aircraft, or ground-aircraft related, programs go to DCMA Instructions 8210.1 and 8210.2 (Reference (dd)) or the DCMA Website at http:/guidebook.dcma.mil/

4.5. Test Operations

The actual execution of test events presents numerous contractor/Government detail-type issues that must be addressed to successfully complete the program and the contract. The following items are potential conflict areas and should be addressed early to ensure clarity and completeness as to contractor and Government responsibilities and expectations for the T&E effort throughout the acquisition process. These areas may or may not be specifically spelled out in the contract but should have been considered during the preparation of the SOW/PWS in some manner.

4.5.1. Change Management

Change is inevitable in any test program. Changes to product/system performance criteria (such as new requirements, deviations, and waivers to existing performance criteria) have to be clearly and completely documented, incorporated into the contract, and adhered to. There should be an approved change management process defining the authority controlling the change process and configuration management of test assets. This is sometimes called a configuration control process, but a distinction needs to be made between the configuration control process that is part of the systems engineering process and is focused on the design configuration and the configuration control process that is focused on test asset configuration. The latter process will include design changes in addition to deviations or waivers resulting from the production process, and even changes to the test instrumentation. The integrity of the test results rests on understanding and maintaining control of the configuration of the test assets as the test program progresses. Unknown or undocumented configuration changes can invalidate data and introduce safety risks. This is especially true with software changes. For more specifics on this topic, see Part 48 of Reference (e).

4.5.2. Reporting

This area requires a very clear contractual understanding and specifics that identify the type, format, schedule, and approving and coordinating authorities for all T&E reports. The contractor is obligated to deliver only the reports listed in the contract as CDRL.

4.5.3. T&E Team Responsibilities

The contract defines the responsibilities of the contractor versus the Government. However, the contract should not be expected to address all of the roles and responsibility issues that arise during the test program execution. It is the responsibility of all parties, but especially the Government representatives, to understand the roles, authority, and span of control of each of the team representatives. The contractor is required only to execute the contract and is not required to do anything above that minimum requirement. A contractor with total system performance responsibility is also responsible for any interface issues that may arise and should have responsibility for identifying any interface issues that may arise involving other contractors or with GFE or supplies. Otherwise, the issue of responsibility for addressing interface issues will need to be worked out on an ad hoc basis. The KO should be consulted to ensure that no constructive changes are incurred to the contract.

Other common issue areas include providing people, spares, and consumables. The responsibilities for data authentication and data access also need to be addressed. *Who will capture the raw data and convert it into useful data products?* If the contractor is responsible for first-generation data processing (data authentication process), will the contractor be responsible only for the data that it intends to analyze, or will the contractor be responsible for processing all data and providing that data to the appropriate Government or contractor for analysis and evaluation?

The contractor may interpret its responsibility as providing data authentication services only for specification compliance-related data, whereas the Government may have assumed that the contractor would provide authentication for all data. In this case, it may help to make it clear that even though the contractor will have to provide data authentication services for all test participants, it will be responsible for analyzing only the data necessary to show compliance with the contract.

4.5.4. Test Personnel

Because contractor and Government personnel work closely together during the execution of test events, it is important to have a clear understanding of what each party is providing in terms of personnel and how they will be managed. The skill sets needed for executing the program need to be identified before the start of the test program. Depending on the product/system under test, there may be a requirement for some specific skill sets to fully exercise the product/system. Once the personnel requirement is established, the source of the personnel should be clearly established. For example, which skills will the contractor acquire for the test program, or from the Government? In some programs, the contractor brings the test managers and the Government provides the maintenance personnel. Whatever the actual arrangement is between contractor- and Government-supplied personnel, clear expectations need to be set as to skill sets and quantity of personnel.

In addition, the contractor and Government management roles and responsibilities must be clear. Do contractor personnel oversee operations with Government personnel? If so, what are the rules governing such issues as work-hour expectations and disputes?

Do Government personnel oversee contractor personnel? If so, how does the government keep from unintentionally making constructive changes to the contract? See subsection 1.3.16. of this guide for discussion of contractors supporting OT&E. Consequently, system contractor personnel may not participate in data authentication groups or RAM scoring conferences or act as data collectors, reducers, or processors.

4.5.5. T&E Team Participants and Roles

The participants in the T&E team include anyone and everyone necessary to successfully execute the test program, or anyone with a stake in the outcome of the test program. Different acquisition programs may have several teams working on T&E issues, but the basic issues to be addressed are management and execution.

The T&E WIPT is generally the team that addresses the approach and overall management of the T&E program, whereas a test team, or something similar, will handle the execution of the test program. The T&E WIPT will include all stakeholders for the approach and status of T&E. At a minimum, T&E WIPT participants include the PM and staff representatives, oversight organizations, contractor and major subcontractors, the responsible test organization, OTA/system evaluator, and appropriate user representatives.

T&E WIPT participants will provide the day-to-day management, execution, and logistics support necessary to plan, execute, analyze data, and report test results. T&E WIPT participants represent different levels of management and perhaps different detailed objectives, so good team management skills will be necessary to establish common goals, minimize conflict between team participants, and execute a timely, efficient, and effective T&E program.

4.5.6. Test Safety and Environmental Issues

The actual testing of equipment in a lab or on a test range introduces personnel safety issues and concerns. For example, the F-16 used hydrazine, a toxic chemical, for its emergency power unit. When the emergency power unit was tested on the ground, ground personnel near the aircraft were exposed to a potentially hazardous environment from hydrazine in the power unit exhaust, and when hydrazine was spilled during servicing of the aircraft, the safety-related aspects were not clear in terms of how to clean up the spill, safe exposure levels, etc. This example illustrates that Government and contractor roles and responsibilities for the conduct and approval of test-related safety issues and analyses need to be clearly defined. Note that in addition to safety analyses for personnel and test article risks, these analyses should also address environmental impacts related to the conduct of tests. Some of these environmental issues are at the State and local level, and the complete list of environmental laws may not be known prior to contract award. As a result, the contract needs to allow for these types of analyses and impacts on the execution of the test program. See section 4.47.1 of Reference (b). Reference (a) requires the PM to certify the safety of the system before Government personnel operate a system under test.

4.5.7. Risk Acceptance Authority

The conduct of safety analyses will assist in identifying and clarifying the risks involved in the test program. Detailed test planning should establish test conditions and test procedures to mitigate most of the significant risks. However, some residual risk will remain, and the question then becomes one of who has the authority to accept the residual risk and allow the test to proceed. The approval authority can be different depending upon the levels of risk established (e.g., low, medium, or high risk).

For example, most flight tests involve a routine or relatively low level of residual risk, so the operations officer or the test team lead has the authority to approve a flight with that level of risk. However, flight tests such as high angle-of-attack (or stall) testing are usually considered to be high-risk tests because the aircraft behavior in the stall regimen is not well known, and the risk of losing the aircraft is considerable. In this case, the range commander or equivalent would be the approval authority for that particular test event. Because the approval (or lack of approval) to conduct tests is not within the contractor's control, the contract needs to account for that possibility. DoD Components may use different risk matrixes, such as 3-tier versus 4-tier or dollar/injury/mission impact thresholds. These different matrixes may also have their own risk decision authority levels. This becomes very important when contracting for an integrated testing program that will cross DT and OT lines, as well as multi-Service OT&Es.

4.5.8. Accident/Incident Investigation and Reporting

In the unfortunate event of an accident or incident, the accident/incident reporting and investigation procedures and process must be clearly defined. This process should include authority, documentation, and accountability for the test article in case of an accident/incident. For example, if a test aircraft crashes, who is going to be held responsible for that test article? Will the accident investigation be conducted according to Government procedures or contractor procedures? How is the contractor expected to support the accident investigation? Will the Government indemnify the contractor for the loss of the test asset, or is the contractor expected to procure insurance to cover the risk of losing the test asset?

4.5.9. Detailed Test Planning

This area refers to detailed test plans or the test plans that are actually constructed and used to execute the test events and acquire the necessary data. Higher levels of test planning, such as T&E strategies and system-level test plans, have more of a management focus and are not sufficiently detailed to actually execute a test event. Where actual test operations are concerned, the detailed test plans drive the actions of the testers. Therefore, the roles and responsibilities for the development of detailed test plans must be defined. This area includes processes for detailed test planning, especially with integrated testing; test plan authorship; and test plan approval. A key consideration is as follows: When the contractor writes the detailed test plans, how does the Government ensure that the contractor does not become responsible for doing more testing than is required for the contract? This issue is part of defining the Government's role in approving detailed test plans.

4.6. Test Execution

The roles and responsibilities for the actual conduct of a test must be defined; that is, essentially, to define who controls the conduct of tests – Government or contractor, or both. This area includes such items as deleting or adding test points, expectations for a particular priority when it comes to range or range asset availability, and contractor or Government run-through of the data collection instrumentation prior to the actual test to verify operational status.

4.6.1. Test Data Access, Authentication, and Sharing

The access to, process for authentication, and sharing of all test data must be clearly established. According to paragraph 2.c.(7) of Enclosure 6 of Reference (a), the government shall have full and timely access to all available developmental, operational, and live-fire T&E data, records, and reports. Government access to all test data should not be restricted, and the process to authenticate test data should be agreed upon. The contract should clearly describe the collection, authentication, and availability process. If a data authentication group is established, the contract must define who will be the leader, where the data will be stored, and how the authenticated data will be made available for all stakeholders. This is an area that will potentially invoke contractor intellectual property issues, so that part of the contract needs to be clearly understood by the test team.

4.6.2. Test Data Analysis and Evaluation

Data analysis and evaluation responsibilities, process, and products must be identified and adhered to throughout the testing effort. The process should clearly identify what the contractor is responsible for, as opposed to the Government, and the process for adjudicating conflicting evaluations. Especially in the case of integrated testing, considerable data will be collected. The contractor should be responsible for analyzing only that data that is sufficient to demonstrate compliance with the specification and SOW/PWS. This area requires a very clear contractual understanding and specifics to identify the type, format, schedule, and approving and coordinating authorities for all T&E reports. The required contractor reports should be listed as contract deliverables. For example, if the Government is expecting or relying on a contractor report to satisfy an acquisition milestone or decision review, then that information needs to be communicated to the contractor.

5. SUMMARY

This guide provides the major T&E items or requirements to consider as T&E professionals develop or review a SOO, SOW, PWS, RFP, and contract. The various lists provide a baseline for discussions, decisions, and review for T&E items or requirements. These lists, combined with DoD Component-specific T&E direction, guidance, and requirements, should help developers and reviewers address all the necessary T&E contents for a SOO, SOW, PWS, and RFP for a program.

The key issue to remember: If a T&E item or requirement is not in the SOW/PWS, it probably will not be in the RFP. If it is not in the RFP, it probably will not be in the contract. If it is not in the contract, *do not expect it!*

T&E professionals must be involved early and stay involved with the PM, the KO, the SE, and the other program office leads throughout the contracting process to ensure that the T&E policies, practices, procedures, and requirements are understood, accepted, and included in the contract as necessary for program success.

APPENDIX A

GETTING STARTED WITH DEFENSE TRADE

THE DIRECTORATE OF DEFENSE TRADE CONTROLS (DDTC) AND THE DEFENSE TRADE FUNCTION

Contents:

I. Does Defense Export Controls Apply to Me? A Quick Action Checklist

- Find out if what you want to export (hardware, technical data, and/or defense services) is covered in the U.S. Munitions List (USML), found in Part 121 of the ITAR.
- Not sure if your desired export is covered by the USML? File a Commodity Jurisdiction request.
- If what you want to export is on the USML, you must be registered with DDTC.
- After you are registered, you may apply for an export license. D-Trade is the preferred way of licensing.
- Have basic questions you need answered? Call the DDTC Response Team.

II. Rationale for Regulating Defense Exports

The U.S. Government views the sale, export, and re-transfer of defense articles and defense services as an integral part of safeguarding U.S. national security and furthering U.S. foreign policy objectives. Authorizations to transfer defense articles and provide defense services, if applied judiciously, can help meet the legitimate needs of friendly countries, deter aggression, foster regional stability, and promote the peaceful resolution of disputes. The U.S., however, is cognizant of the potentially adverse consequences of indiscriminate arms transfers and, therefore, strictly regulates exports and re-exports of defense items and technologies to protect its national interests and those interests in peace and security of the broader international community.

III. DDTC – The Offices that Administer the Defense Export Regulations

The Directorate of Defense Trade Controls (DDTC), Bureau of Political-Military Affairs, in accordance with 22 U.S.C. 2778-2780 of the Arms Export Control Act (AECA) and the International Traffic in Arms Regulations (ITAR) (22 CFR Parts 120-130), is charged with controlling the export and temporary import of defense articles and defense services covered by the United States Munitions List (USML). To learn more about DDTC, please visit its Web site (www.pmddtc.state.gov).

IV. Authority for Control of Arms Exports (AECA)

The AECA provides the authority to control the export of defense articles and defense services. The AECA charges the President to exercise this authority, which has been delegated to the Secretary of State. The AECA is available through the DDTC Web site. The ITAR implements the AECA. These regulations are frequently updated and revised to reflect change in the international political and security climate, as well as technological development. The ITAR may be accessed on the DDTC Web site.

In accordance with Executive Order 11958, the State Department, with the concurrence of the Department of Defense, determines what commodities are covered by the USML. Guidance on the commodity jurisdiction (CJ) function is available on the DDTC Web site. In addition to seeking technical support and national security assessments from the Department of Defense, the State Department relies on extensive interagency cooperation and coordination to perform the arms export control function. It:

- Works closely with U.S. Customs and Border Protection (review of defense industry registration, performance of defense export end-use checks, investigations, civil penalties);
- Works with the Intelligence Community to review alleged diversions and unauthorized transfers; and
- Cooperates with the Justice Department and U.S. Attorneys (pre-trial consultations, trial documentary preparation, expert testimony).

V. U.S. Government Regulatory Measures

For the U.S., licensing and compliance are two sides of the same coin, and there is constant interaction between the two functions.

I. Registration

- In accordance with the AECA, registration with the State Department (via DDTC) of all U.S. persons that manufacture or export defense articles, furnish defense services, or U.S. and foreign persons engaged in arms brokering, is required. The information submitted by registrants is reviewed by the Treasury Department to ensure there are no outstanding law enforcement concerns.
- Registration does not confer any export privileges, but is a prerequisite to export licensing approval.
- The registration process:
  - o Informs the U.S. Government about the U.S. defense industry (legal status, export eligibility, foreign ownership/affiliations, legally responsible personnel, areas of activity);
  - o Serves as a channel to provide industry with information about export regulations and Government concerns; and
  - o Helps validate the bona fides of U.S. firms engaged in defense trade, especially during the review of export license applications.
- Registrants, in accordance with the AECA, are charged a fee. Congress has created a mechanism that allows the State Department to retain the money collected to help support defense export control functions.

To learn more about registration, and to access the registration form, please visit the "Registration" page on the DDTC Web site.

II. Licensing

- Department of State approval of a license application is required prior to the export of defense articles or defense services.
- About 30 percent of the applications processed by the State Department are referred to other offices and agencies (e.g., the Department of Defense) for comment and recommendations. This is what is referred to as “staffing” the case.
- During the review process, a computerized review of all parties to the proposed transactions is made against a “watch list” of known or suspected export violators. A “match” results in a full compliance review by the State Department before final action is taken on the application.
- In addition to sorting through detailed technical specifications, the license application review process clarifies the ultimate end-use and end-user of the defense export, as well as facts related to intermediate handling.
- From the enforcement point of view, the review process provides an avenue to prevent or eliminate diversions, and to assist the U.S. Government in investigations and prosecutions should an export violation be suspected or reported.
- In submitting license applications, companies must certify eligibility to export and an understanding of the laws governing such exports. Moreover, in carrying out the physical act of exporting, they must meet certain conditions in terms of documentation (electronic reporting of export information using the Automated Export System, “AES”) and handling (particularly of classified material).
- Exporters must make clear on shipping documents that the defense export cannot be resold or retransferred without prior U.S. Government authorization – a licensing requirement that also involves compliance issues.

To learn more about licensing, and to view the various licensing forms, please visit the D-Trade Info Center and the “Licensing” page on DDTC’s Web site.

VI. End-Use/End-User Monitoring

End-Use checks are key to the State Department’s effort to prevent illegal defense exports and technology transfers.

End-use checks (known under the program name “Blue Lantern”) enlist the help of U.S. diplomatic posts, the cooperation of U.S. Customs and Border Protection, and, most importantly, foreign governments in the conduct of pre-license checks and post-shipment verifications of defense exports.

End-Use monitoring reports are available on DDTC’s Web site.

VII. Other Compliance Mechanisms

The U.S. Government spends considerable effort trying to prevent violations, via participation in industry conferences, Internet postings, and publication of regulations. When a problem arises, the Department of State has broad authority to take action (i.e., suspend, deny, or revoke license approvals). Working with law enforcement agencies, it can prosecute criminally (possible prison sentences and fines) and independently can take civil action (e.g., fines and denial of export privileges).

Remedial assistance/attention is also offered. The State Department works with companies to develop effective export compliance programs. DDTC makes available a guideline describing the basic elements of a compliance program via its Web site.

VIII. D-Trade – Conducting Your Defense Trade Business Electronically

Effective January 15, 2004, DDTC, through the use of the D-Trade electronic licensing system, is prepared to receive and adjudicate fully electronic defense export authorization requests properly submitted by any U.S. person who is a defense trade registrant and wishes to permanently export unclassified defense articles via the Form DSP-5, temporarily import unclassified defense articles via the Form DSP-61, or temporarily export unclassified defense articles via the Form DSP-73. Based on envisioned expansion of electronic processing capabilities, DDTC anticipates, with few exceptions, most export licensing submissions via D-Trade in the near future.

For more information on D-Trade, consult the D-Trade Information Center, accessed through the DDTC home page. There you will find links to more background information on electronic licensing.

IX. To Learn More

The DDTC Web site has more information that may be useful to you. The homepage has a comprehensive listing of links to information that can assist you in your defense exporting endeavors; in addition, consult "New Items and Announcements" for the latest updates.
If you have any questions about any aspect of the defense export process, please contact the DDTC Response Team.

APPENDIX B

SAMPLE CHECKLIST FOR EVALUATING AN RPP

This checklist (adapted from Reference (n)) is not meant to be all inclusive but rather serves as a tool to guide discussions and decisions relative to RAM planning, accountability, and reporting for your programs.

RPP

Does the program…

- Implement the reliability activities described within the RPP with appropriate methods, tools, and best practices, in order to accomplish the following four objectives: understand the Government's requirements, design product/system for reliability, produce reliable products/systems, monitor and assess user reliability?
- Include a reliability growth plan?
- Include procedures for verifying that planned reliability activities are implemented?
- Manage risks due to new technologies?
- Include decision-making criteria and plans for intensifying reliability-improvement efforts?
- Require periodic updates coordinated with the customer/user?

System Reliability Model

Does the program…

- Build and refine model throughout the life cycle?
- Routinely update the model as failure definitions are updated, failure modes are identified, operational and environmental load estimates are updated, and design or manufacturing changes are made?
- Include detailed component stress and damage models?
- Use the model to (1) update allocations, (2) aggregate reliability, (3) identify single points of failure, and (4) identify critical reliability items and the need for additional design or testing activities?

Systems-Engineering Integration

Does the program…

- Integrate reliability activities with the systems engineering process throughout the life cycle?
- Incorporate reliability improvement actions routinely during design, production, and in the field?
- Monitor and evaluate the reliability impact of design changes and supplier change notices throughout the life cycle?
- Manage and control critical reliability items?
- Adhere to design rules that affect reliability?

System-Level Operational and Environmental Life Cycle Loads

Does the program…

- Develop and periodically update load estimates throughout the life cycle?
- Verify estimates on instrumented systems/products with operationally realistic conditions applied in time for reliability verification?
- Use estimates in reliability modeling, assessment, and verification?
- Coordinate estimates with the systems engineer?

Life Cycle Loads on Subsystems, Assemblies, Subassemblies, and Components

Does the program…

- Develop and periodically update these load estimates based on operational and environmental loads applied at the system level?
- Verify load estimates on instrumented systems/products/assemblies with operationally realistic conditions applied?
- Flow down estimates and updates to designers; integrators of COTS, NDI, and GFE; and suppliers?
- Use estimates to identify failure modes and mechanisms and in assessments and verification?

Identify and Characterize Failure Modes and Mechanisms

Does the program…

- Identify failure modes and mechanisms throughout the life cycle?
- Begin to identify failure modes and mechanisms as soon as development begins using realistic life cycle operational and environmental loads in conjunction with engineering- and physics-based models?
- Ensure that teams developing assemblies, subassemblies, and components for the system identify and confirm failure modes and distributions with analysis, test, or accelerated test?
- Ensure that teams selecting/integrating assemblies, subassemblies, and components for the system (including COTS, NDI, and GFE) identify and confirm failure modes and distributions with analysis, test, or accelerated test?
- Identify and confirm failure modes induced by manufacturing variation and errors?
- Identify and confirm failure modes induced by user or maintainer errors?
- Analyze all test and field failures to root cause?

Closed-Loop Failure-Mode Mitigation

Does the program…

- Analyze and map to the customer-specified failure definition and scoring criteria (FDSC) for all failure modes to formulate corrective actions throughout the life cycle?
- Aggressively mitigate failure modes until reliability requirements are met?

- Employ a mechanism for monitoring and communicating the implementation and effectiveness of corrective actions that is accessible by the customer?
- Include failure modes that may occur during the life cycle in the system reliability model?

Reliability Assessment

Does the program…

- Assess feasibility of reliability requirements using the system reliability model in conjunction with expert judgment?
- Allocate reliability requirements to lower indenture levels and flow them to subcontractors/suppliers?
- Periodically assess reliability of the system throughout the life cycle using the reliability model, the life cycle operational and environmental load estimates, and the customer-specified FDSC?
- Include reliability values to be achieved at various points in the program?
- Track reliability assessments from analysis, M&S, test, and the field as a function of time and compare them with allocations and customer reliability requirements?
- Monitor and evaluate the implementation of corrective actions as well as other changes to the design or manufacture of the systems/product that may impact reliability?
- Include COTS, NDI, and GFE in all assessments?

Reliability Verification

Does the program…

- Develop and periodically refine a reliability requirements verification strategy/plan that is an integral part of the systems engineering verification and is coordinated and integrated across all phases?
- Include a strategy to ensure that reliability requirements will be verified during design and will not degrade during production or in the field?
- Include in a reliability growth plan the reliability values to be achieved at various points during development?
- Base verification on analysis, M&S, testing, or a mixture, and ensure that the verification is operationally realistic?
- Verify that system-level operational and environmental life cycle loads will be used?
- Include any customer-specific requirements?

Failure Definitions

Does the program…

- Understand customer-specified FDSC?
- Design to avoid failures due to user or maintainer errors?
- Ensure that the RPP integrates customer-specified FDSC with (1) system reliability model, (2) identification of failure modes and mechanisms, (3) closed-loop failure-

mitigation process, (4) reliability assessment, and (5) reliability verification throughout life cycle?

Technical Reviews

Does the program…

- Ensure that the RPP specifies how and when technical reviews will be conducted throughout the life cycle?
- Conduct periodic interchanges with the customer/user that promote understanding of operational environment?
- Schedule and conduct technical reviews to (1) ensure progress toward achieving reliability requirements, (2) verify that planned reliability activities are implemented, and (3) compare status and outcomes of reliability activities?
- Ensure that SMEs conduct independent peer review?
- Conduct and participate in reviews with the customer/user that address identification, analysis, classification, and mitigation of failure modes?

Methods and Tools

Does the program…

- Implement reliability activities with methods and tools from the RPP?
- Implement and adhere to reliability best practices?
- Obtain customer approval for changes in methods, tools, or best practices and include these in the RPP?

Outputs and Documentation

Does the program…

- Document a plan for RPP updates?
- Ensure continuous customer access to status and outputs from all reliability activities?
- Schedule and document outputs appropriately in the reliability case?

APPENDIX C

SAMPLE T&E AWARD FEE CRITERIA

This matrix is not meant to be all inclusive but rather serves as a sample to guide discussions and decisions relative to award fee planning, accountability, and program reporting. To the extent that T&E measures of contractor performance can be objectively measured, an incentive fee, rather than an award fee, should be used to motivate excellent contractor performance. Although the samples below may be useful, fee determination must be done solely in accordance with the applicable contract clauses and award fee plan.

EXCELLENT	VERY GOOD	SATISFACTORY	UNSATISFACTORY
T&E reviews met all the entry, exit, and success criteria (including teammates, vendors, and subcontractor reviews). Reviews were successful. Program proceeded as planned. Reliability growth covered with complete risk assessment* on all critical areas.	T&E reviews met most of the entry, exit, and success criteria. Only minor omissions. Reviews were successful although there were minor re-reviews but no significant delays to subsequent events. Reliability growth covered with some risk assessment* provided on most critical areas.	T&E reviews met most of the entry, exit, and success criteria. Reviews were successful although a few items required subsequent re-review. Is consistent with the TES and TEMP, as appropriate, and the SEP. Program experienced some rework with no program impacts to the critical path. Reliability growth covered with risk assessment* provided on some critical areas.	T&E reviews did not meet some of the entry and exit criteria. Omissions are considered significant. Is not consistent with SEP, TES, TEMP as appropriate. Subsequent re-reviews required. Program delays and cost increases experienced. Critical path was affected. Reliability growth not mentioned.
T&E baseline data package is complete with no TBDs, omissions, or incorrect data. Requirements management process is actively used with minimal change rate, no technical discrepancies, and only a few administrative discrepancies. Baselines established ahead of schedule.	T&E baseline data package is mature and stable with only minor TBDs, omissions, or incorrect data. Requirements management process is in place and used with acceptable change rate with only minor technical discrepancies. Baselines established on schedule.	T&E baseline data package is well defined, mostly mature, and stable with no serious TBDs, omissions, or incorrect data. Requirements management process is in place and used with acceptable change rate and no serious technical discrepancies. Baselines established on schedule.	T&E baseline data package only partially defined. Requirements management process experiences high change rate and is in the state of flux. Program delays or cost increases incurred. Critical path is affected.
T&E reflects best practices. Best practices are flowed down to subcontractors, teammates, and vendors. Program execution applies the documented program processes.	T&E reflects best practices and program-specific needs. Best practices are flowed down to principal subcontractors, vendors, and teammates. Program execution applies critical documented program processes.	T&E reflects best practices that are critical to high-risk program areas. Best practices are flowed down to critical subcontractors, vendors, and teammates. Program execution usually applies the documented program processes.	T&E reflects best practices. Best practices are not flowed down to critical subcontractors, vendors, and teammates. Program has deviated from the documented program processes.
Critical path is defined and actively managed. Proactive risk management processes applied across the program to include subcontractors, vendors, teammates, and Government participants. Risk mitigation plans are in place and on schedule.	Critical path is defined and managed. Risk management process includes subcontractors, vendors, teammates, and Government participants. Risk mitigation plans are in place and incorporated into the program. Only minor delays to risk mitigation schedules.	Critical path is defined and managed. Risk management process includes critical subcontractors, vendors, and teammates. Risk mitigation plans are focused on critical path and incorporated into the program. Occasional modification of or addition of risk mitigation plans is needed.	Critical path is ill-defined and not well managed. RMPs are not well defined and do not include the subcontractors, vendors, or teammates. Continual modification of or addition of risk mitigation plans that affect the critical path are needed.
A DR process is clearly identified and part of the review process.	A DR process is in place and is sporadically used in reviews.	A DR process is in place but not regularly used.	A DR process is in place but not used.

*See the risk assessment matrix in Appendix F.

APPENDIX D

SAMPLE PAST PERFORMANCE QUESTIONNAIRE

This questionnaire is not meant to be all-inclusive; instead, it serves as a tool to guide discussions and decisions regarding ranking contractor past performance relative to the program. Although the samples below may be useful, evaluation of proposals must be done solely in accordance with the applicable SSP and RFP evaluation factors.

Sample Past Performance Questionnaire
Based on your knowledge of the contract, please provide your assessment of how well the contractor performed on each of the following topics. Only performance in the past 5years is relevant. Please check the appropriate rating and comment on all responses other than those rated Satisfactory or N/A.

Performance Rating Definitions:

Exceptional (E)	Very Good (V)	Satisfactory (S)	Marginal (M)	Unsatisfactory (U)	N/A
Indicates performance clearly exceeded requirements. Area of evaluation contains few minor problems for which corrective action appears highly effective.	Indicates performance exceeded some requirements. Area of evaluation contains few minor problems for which corrective action appears effective.	Indicates performance meets contractual requirements. Area of evaluation contains some minor problems for which the corrective actions appear satisfactory.	Indicates performance meets contractual requirements. Area of evaluation contains a serious problem for which corrective actions have not yet been identified, appear only marginally effective, or have not been fully implemented.	Indicates the contractor is in danger of not being able to satisfy contractual requirements and recovery is not likely in a timely manner. Area of evaluation contains serious problems for which the corrective actions appear ineffective.	Neutral or Unknown

Sample Questions

Was there a single T&E authority designated for the program with clear lines of authority and responsibility to the PM?	E	V	S	M	U	N/A
Did the contractor include Government T&E personnel on the IPTs to create an integrated team approach?	E	V	S	M	U	N/A
How well did the contractor maintain a balanced set of system performance, cost, and schedule requirements during the program?	E	V	S	M	U	N/A
Did the contractor use its best practice software development process work across the total industry team?	E	V	S	M	U	N/A
How effective was the contractor's interface management and control?	E	V	S	M	U	N/A
How well did the contractor manage technical risk? Was it focused on the risks associated with the critical path?	E	V	S	M	U	N/A
Did the contractor complete all theT&E entry/exit criteria for major design reviews effectively? Were action items established and expeditiously closed?	E	V	S	M	U	N/A
Did the contractor deliver quality T&E products (reports, analyses, trade studies, LDTs, and specifications) in a timely mannor?	E	V	S	M	U	N/A
How well did the contractor manage event-based reviews with its subcontractors, teammates, and vendors?	E	V	S	M	U	N/A
Did the contractor include SMEs in T&E reviews on higher-risk areas of the program?	E	V	S	M	U	N/A
Did the contractor apply the corporate best T&E practices and use its experienced personnel?	E	V	S	M	U	N/A
How well did the contractor adhere to the program T&E schedule in the execution of the program?	E	V	S	M	U	N/A
How well did the contractor maintain the program T&E process? Was it updated with the results of CPI efforts?	E	V	S	M	U	N/A
Were the T&E requirements extended to subcontractors, teammates, and vendors?	E	V	S	M	U	N/A
How well did the contractor integrate theT&E processes and tools in the management of the program (SEP, IMP, IMS, EVM, risk management)?	E	V	S	M	U	N/A
How well did the contractor manage the performance baselines of the program?	E	V	S	M	U	N/A
How well did the contractor employ metrics (e.g., deliquency reporting, reliability growth) to manage performance baseline maturity?	E	V	S	M	U	N/A
How timely, complete, and usable was the T&E data package for the defined performance baselines? Was the T&E data package complete to support the program's technical and acquisition strategy?	E	V	S	M	U	N/A
How well did the contractor manage the requirements and apply any requirements management tool? Did the program experience an unusually high requirements change rate?	E	V	S	M	U	N/A

APPENDIX E

AREAS WITHIN THE RFP FOR DT&E FOCUS

RFP Section	Title	Description	Information to Review Within Section
B	Supplies or Services and Process/Costs	This section includes a brief description of the supplies or services (e.g., item number; national stock number; part number, if applicable; nouns, nomenclature, and quantities) and includes incidental deliveries such as manuals and reports.	Review all CDRLs for inclusion of T&E execution support (i.e., data rights, test data, test plans, source code drop, prototype quantity, delivery times/location).
C	Description/ Specifications/ SOW	This section contains detailed description of the products to be delivered or work to be performed under the contract and the preliminary SPS.	Are all requirements clearly defined and stated in performance-based terms?
			Are performance-based characteristics directly tied to program objectives?
E	Inspection and Acceptance	This section includes inspection, acceptance, quality assurance, and reliability requirements.	Has the acquisition team developed a tailored quality assurance surveillance plan to monitor contractor performance?
			Describe the organization and procedures to perform R&M task.
F	Deliveries of Performance	KO will specify the requirements for time, place, and method of delivery of performance.	Has the required number (sample size) of test articles been identified?
			Has a delivery location along with schedule been established?
			If you think you may want the contractor-acquired property, have the KO state in the solicitation and resulting contract that title to the contractor-acquired property will revert to the Government at the end of the contract.

RFP Section	Title	Description	Information to Review Within Section
			Identify PM's desire to have contractor support personnel available to repair or provide reach-back of contractor's product during DT&E effort.
			Identify contractor property needed as spares during the testing.
H	Special Contract Requirements	KO will include statement of any special contract requirements that are not included in Section I, Contract Clauses, or in other sections of the uniform contract format.	Review all contract clauses for data delivery, Government property, rent-free Government property, and personnel qualifications. This information may reside in Section H or I or both sections.
I	Contract Clauses	KO shall include in this section the clauses required by law or by any additional clauses expected to be included in any resulting contract, if these clauses are not required in any other section within the uniform contract format. An index may be inserted if this section's format is particularly complex.	Review all contract clauses for data delivery, Government property, rent-free Government property, and personnel qualifications. This information may reside within Section H or I or both sections.
J	List of Attachments	KO shall list the title, date, and number of pages for each attached document, exhibit, and other attachment. Cross-references to material in	Identify whether TEMP is releasable to the contractor. If so, make sure it is provided to the contractor and listed within Section J, List of Attachments.

RFP Section	Title	Description	Information to Review Within Section
		other sections may be inserted as appropriate.	Documents released to contractor should be reviewed for security classification. Those documents non-releasable to the public should have a distribution list established so they can be viewed by the companies performing the work.
K	Representations, Certifications, and Other Statements of Offerors or Respondents	Include in this section those solicitations provisions that require representations, certifications, or the submission of other information by offerors.	Review for requests for certain certifications that support T&E strategy.
L	Instructions, Conditions, and Notices to Offerors or Respondents	Insert within this section of solicitation the provisions and other information and instructions not required elsewhere to guide offerors or respondents in preparing proposals or responses to RFIs. Prospective offerors or respondents may be instructed to submit proposals or information in a specific format or several parts to facilitate evaluation. The instructions may specify further organization of proposal or response parts, such as administrative, management, technical, past performance, and certified cost of pricing data.	If contractor will provide oversight for another contractor or direct work to another contractor, what measures are planned or have been taken to reduce or eliminate potential organizational conflicts of interest?
			Describe the contractor test management structure for T&E, experience of T&E staff, the predicted staffing levels, and the application of T&E best practices.
			Define the responsibilities of the contractor and the Government during test planning (include contractor testing, DT, and integrated testing).
			Describe contractor's approach on technical data, including management, ownership, control, timely access, and delivery of T&E data, to include raw test

RFP Section	Title	Description	Information to Review Within Section
			data, to support the evolving technical baseline.
			Define CDRLs and DIDs. Identify any T&E related data products that contractor must provide. Determine applicability of DIDs in support of T&E efforts.
			Determine applicability of commercial certifications of material or product.
			Does the RFP contain a top-level schedule depicting key T&E events?
			Are the allocation of M&S responsibilities between the Government and contractor and the expectations regarding the sources of M&S tools described?
			Has the acquisition team identified an industry day?
			Define release of T&E assessment data to industry.
			Is the program or aspects of said program classified? If so, is contractor capable of storing, handling, obtaining, and controlling classified data? Are contractor T&E personnel cleared to review program? Contractor should provide certification of classification capability along with designated personnel.
			Is the acquisition team providing a copy of or access to the program TEMP or T&E strategy?

RFP Section	Title	Description	Information to Review Within Section
			Does the RFP set the expectations (scope, phasing, rapid acquisition authority) of an ITT and associated plan?
			Does the RFP require the bidder to submit a comprehensive facilities plan that includes identification of existing infrastructure (both industry and Government) and bidder investment requirements for expanded or additional infrastructure?
			Does the RFP incentivize the bidder to use COTS, and does it require the bidder to show associated risks and opportunities (is more testing required, or less) with COTS applications?
			Does the RFP explain how interoperability is validated in the program's test phase?
			Is there linkage between the preliminary design specification and the test requirements?
			Do testing requirements in the RFP include procedures for ensuring the pedigree of the data to include government review and approval of contractor test plans prior to execution of test event, government witness of test event, and government review and approval of final test report/analysis?

RFP Section	Title	Description	Information to Review Within Section
M	Evaluation Factors for Award	Identify all significant factors and any significant subfactors that will be considered in awarding the contract and their relative importance.	Has the acquisition team mapped Sections L and M to the program supporting documents (acquisition strategy/plan, TEMP, and SSP) and requirements document (ICD, CDD, or CPD)?
			Are minimum thresholds and maximum performance objectives clearly defined?
			Are requirements stated in certain terms such that evaluators will be able to assess whether the offeror meets or exceeds a particular outcome?
			What are the measures and metrics to evaluate qualification of contractor T&E personnel?
			Are critical program objectives reflected in the evaluation criteria?
			Does contractor propose to use IMP/IMS/EVMS in an integrated manner to verify program entry and exit criteria by developmental phase?
			How does contractor propose to use the verification cross-reference matrix verification method to verify/burn down system-level requirements associated with KPPs/TPMs by program phase?

RFP Section	Title	Description	Information to Review Within Section
			How does contractor propose to compare and use KPP/TPM from COI to CTP measure of effectiveness to measure of suitability? (This is an effort to align this document with DTM 09-027 (Reference (cc) and integrated testing as outlined in the DAG (Reference (b))).
			How would contractor develop and implement a data analysis plan to support "Program" integrated testing?
			Has the contractor shown linkage between the preliminary design specification and the test requirements?

APPENDIX F

RELIABILITY ANALYSIS, PLANNING, TRACKING, AND REPORTING

In accordance with Reference (n), the sustainment characteristics of the materiel solution resulting from the AoA and the CDD sustainment KPP thresholds have been translated into R&M design requirements and contract specifications. The strategies shall also include the tasks and processes to be stated in the RFP that the contractor will be required to employ to demonstrate achievement of reliability design requirements. The TES and the TEMP shall specify how reliability will be tested and evaluated during the associated acquisition phase.

Reliability growth planning should consider the reliability targets (initial and goal), test phases, CAPs and reliability thresholds (interim goals to be achieved following CAPs). Reliability growth planning should also include realistic management metrics, such as management strategy and fix effectiveness factors.

RGCs can be developed using models, such as the Planning Model based on Projection Methodology (PM2). An example of a reliability scorecard can be obtained by sending an e-mail to amsaa.reltools@us.army.mil. The proposed RGC can be assessed using the following RGC risk assessment matrix. This matrix defines risk thresholds against each of the key reliability growth parameters.

RGC Risk Assessment Matrix

Category	Low Risk	Medium Risk	High Risk
Mean Time Between Failures (MTBF) Goal (DT)	Less than 70% of Growth Potential	70%–80% of Growth Potential	Greater than 80% of Growth Potential
IOT&E Producer's Risk	20% or less	20%–40%	Greater than 40%
IOT&E Consumer's Risk	20% or less	20%–30%	Greater than 30%
Management Strategy	90% or lower	90%–96%	Greater than 96%
Fix Effectiveness Factor	70% or lower	70%–80%	Greater than 80%
MTBF Goal (DT)/ MTBF Initial	Less than 2	2–4	4 or Larger
Time to Incorporate and Validate Fixes in IOT&E Units Prior to Test	Adequate time and resources to have fixes implemented and verified with testing or strong engineering analysis	Time and resources for almost all fixes to be implemented and most verified with testing or strong engineering analysis	Many fixes not in place by IOT&E and limited fix verification

Category	Low Risk	Medium Risk	High Risk
CAPs	5 or more CAPs that contain adequate calendar time to implement fixes prior to major milestones	3 or 4 CAPs, but some may not provide adequate calendar time to cut in all fixes	1–2 CAPs of limited duration
Reliability Increases after CAPs	Moderate reliability increases after each CAP result in lower-risk curve that meets goals	Some CAPs show large jumps in reliability that may not be realized because of program constraints	Majority of reliability growth tied to one or a couple of very large jumps in the RGC
Percent of Initial Problem Mode Failure Intensity Surfaced	Growth appears reasonable (i.e., a small number of problem modes surfaced over the growth test do not constitute a large fraction of the initial problem mode failure intensity)	Growth appears somewhat inflated in that a small number of the problem modes surfaced constitute a moderately large fraction of the initial problem mode failure intensity	Growth appears artificially high with a small number of problem modes comprising a large fraction of the initial problem mode failure intensity
Test Phase Length	Sufficient to surface at least 5 problem modes in time to address and fix in test phase CAP	Sufficient to surface 3 or 4 problem modes to result in significant reliability growth after CAP	Only enough to surface less than 3 problem modes to analyze and correct, resulting in small jump in reliability after CAP

A key to reliability program success is that there needs to be a reliability focus on best DfR activities during the design of the system. A DfR program shall be articulated in the SEP and captured in the RPP. The DfR program should be executed prior to MS B to ensure that the program achieves its initial reliability targets during early system-level prototype testing. It is pivotal that a very large portion of failure modes are eliminated prior to MS B. There should be only a few remaining significant failure modes post-MS B that need to be addressed as part of the reliability growth program.

APPENDIX G

DAU COURSES AVAILABLE

Provided is a listing of available DAU courses that will assist in further background and knowledge within this area.

CLM 031, Improved Statement of Work (SOW)
CLL 015, Business case Analysis (BCA)
CLC 011, Contracting for the Rest of us
CLM 016, Cost Estimating
CLM 013, Work Breakdown Structure (WBS)
CLB 007, Cost Analysis

ACRONYMS

AoA	analysis of alternatives
APB	acquisition program baseline
AS	acquisition strategy
ASSIST	Acquisition Streamlining and Standardization Information System
BOE	basis of estimate
C&A	certification and accreditation
C4ISR	command, control, communications, computers, intelligence, surveillance, and reconnaissance
CAP	corrective action period
CBA	cost-benefit analysis
CDD	capability development document
CDR	critical design review
CDRL	contract data requirements list
CLIN	contract line item number
CLM	continuous learning module
COI	critical operational issue
CONOPS	concept of operations
COTS	commercial off-the-shelf
CPAR	contractor performance assessment report
CPI	continuous process improvement
CPD	capability production document
CTF	combined T&E task force
CTP	critical technical parameter
CTT	contractor/combined test team
DAG	Defense Acquisition Guidebook
DAU	Defense Acquisition University
DCMA	Defense Contract Management Agency
DFARS	Defense Federal Acquisition Regulation Supplement
DfR	design for reliability
DIACAP	DoD Information Assurance and Certification Accreditation Process
DID	data item description
DoD	Department of Defense
DoDD	Department of Defense Directive
DoDI	Department of Defense Instruction
DOE	design of experiments
DOT&E	Director of Operational Test and Evaluation
DR	deficiency report(ing)
DT	developmental test(ing)
DT&E	developmental test and evaluation
DTM	Directive-Type Memorandum
EMD	engineering and manufacturing development

EVM	earned value management
EVMS	Earned Value Management System
FAR	Federal Acquisition Regulation
FDSC	failure definition and scoring criteria
FedBizOpps	Federal Business Opportunities
FOC	full operational capability
FoS	family of systems
FRACAS	Failure Reporting and Corrective Action System
FRP	full-rate production
GFE	Government-furnished equipment
GFI	Government-furnished information
HSI	human systems integration
HWIL	hardware-in-the-loop
IA	information assurance
IBR	integrated baseline review
ICD	initial capabilities document
IMP	integrated master plan
IMS	integrated master schedule
IOC	initial operational capability
IOT&E	initial operational test and evaluation
IPT	integrated product team
ITAR	International Traffic in Arms Regulation
ITT	integrated test team
JCIDS	Joint Capabilities Integration and Development System
JITC	joint interoperability test command
KO	contracting officer
KPP	key performance parameter
LDT	limited development test
LFT&E	live-fire test and evaluation
LRIP	low-rate initial production
M&S	modeling and simulation
MIL-STD	Military Standard
MOA	memorandum of agreement
MOSA	modular open systems approach
MRTFB	Major Range and Test Facility Base
MS	milestone
MSA	materiel solutions analysis
MTBF	mean time between failures

OT	operational test(ing)
OTA	operational test agency
OT&E	operational test and evaluation
OUSD(AT&L)	Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics
PDR	preliminary design review
PM	program manager
PPIRS	Past Performance Information Retrieval System
PWBS	program work breakdown structure
PWS	performance work statement
R&M	reliability and maintainability
RAM	reliability, availability, and maintainability
RFI	request for information
RFP	request for proposal
RGC	reliability growth curve
RMP	risk management plan
RPP	reliability program plan
SE	systems engineer
SEP	systems engineering plan
SFR	system functional review
SME	subject matter expert
SOO	statement of objectives
SoS	system of systems
SOW	statement of work
SPS	system performance specification
SRR	system requirements review
SSP	source selection plan
T&E	test and evaluation
TBD	to be determined
TD	technology development
TEMP	test and evaluation master plan
TES	test and evaluation strategy (document)
TPM	technical performance measurement
TRL	technology readiness level
TRR	test readiness review
U.S.C.	United States Code
VV&A	verification, validation, and accreditation
WBS	work breakdown structure
WIPT	working integrated product team

REFERENCES

(a) DoD Instruction 5000.02, "Operation of the Defense Acquisition System," December 8, 2008

(b) Defense Acquisition Guidebook, https://dag.dau.mil

(c) Office of the Deputy Under Secretary of Defense for Acquisition and Technology, Systems and Software Engineering, "Guide for Integrating Systems Engineering into DoD Acquisition Contracts," Version 1.0, December 11, 2006

(d) Defense Federal Acquisition Regulation Supplement (DFARS) and Procedures, Guidance, and Information (PGI) http://www.acq.osd.mil/dpap/dars/dfarspgi/current/index.html

(e) Federal Acquisition Regulation (FAR) Website: http://www.acquisition.gov/comp/far/index.html

(f) Office of the Deputy Under Secretary of Defense for Acquisition and Technology, Systems and Software Engineering, "Integrated Master Plan and Integrated Master Schedule Preparation and Use Guide," Version 0.9, October 21, 2005. Available at https://acc.dau.mil/

(g) Defense Acquisition Univeristy (DAU) Acquisition Community Connection (ACC). https://learn.dau.mil

(h) U.S.C. Title 10. Searchable website: http://uscode.house.gov/search/criteria.shtml

(i) Department of the Air Force Award Fee Guide, October 2008. Website: http://www.acq.osd.mil/dpap/ccap/cc/jcchb/Files/Topical/1Restricted/award.fee.oct08.pdf

(j) Army Contracting Agency Award Fee Board Member Guide, Setpember 2003

(k) Department of the Navy, Navy-Marine Corps Award Fee Guide, July 2004

(l) DoD Directive 5000.01, "The Defense Acquisition System," May 12, 2003. Available at http://www.dtic.mil/whs/directives/corres/pdf/500001p.pdf

(m) Chairman of the Joint Chiefs of Staff Instruction (CJCSI) 3170.01G, "Joint Capabilities Integration and Development System,"

(n) Directive-Type Memorandum 11-003, "Reliability Analysis, Planning, Tracking, and Reporting," March 21, 2011

(o) MIL-STD-46855A, "Human Engineering Requirements for Military Systems, Equipment, and Facilities," May 24, 2011

(p) MIL-STD-1472F, "Human Engineering," August 23, 1999

(q) DoD Instruction 8510.01, "DoD Information Assurance Certification and Accreditation Process (DIACAP)," November 28, 2007

(r) Director, Defense Procurement and Acquisition Policy Memorandum, "Proper Use of Award Fee Contracts and Award Fee Provisions," April 24, 2007

(s) Undersecretary of Defense for Acquisition, Technology, and Logistics (USD(AT&L) Memorandum, "Better Buying Power Guidance for Obtaining Greater Efficiency and Productivity in Defense Spending," September 14, 2010

(t) Air Force (SAF/AQC) Award Term/Incentive Options Guide, January 2003 http://www.safaq.hq.af.mil/contracting/toolkit

(u) Deputy Under Secretary of Defense (Acquisition and Technology) Memorandum, "Award Fee Contracts (FAR 16, DFARS 215, DFARS 216)," March 29, 2006

(v) DAU "Acquisition Community Connection Contract Cost, Price, and Finance Website," https://acc.dau.mil/awardandincentivefees

(w) Under Secretary of Defense for Acquisition and Technology Memorandum, "Incentive Strategies for Defense Acquisitions," January 5, 2001

(x) Federal Business Opportunities (FedBizOpps) Website: https://www.fbo.gov

(y) Deputy Under Secretary of Defense for Acquisition and Technology and Director, Operational Test and Evaluation Memorandum, "Definition of Integrated Testing," April 25, 2008

(z) MIL-STD-882D, "Department of Defense Standard Practices for System Safety," February 10, 2000

(aa) DoD Directive 5000.59, "DoD Modeling and Simulation (M&S) Management," August 8, 2007

(bb) MIL-HDBK-881, "Work Breakdown Structure for Defense Materiel Items," July 30, 2005

(cc) DTM 09-27, "Implementation of Weapons System Acquisition Reform Act of 2009"

(dd) Defense Contract Management Agency Instructions 8210.1 and 8210.2 http://guidebook.dcma.mil/

Incorporating Test and Evaluation into Department of Defense Acquisition Contracts

October 2011

Office of the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation

3040 Defense Pentagon
Washington, DC 20301

APPENDIX E – ABBREVIATIONS, ACRONYMS, AND SYMBOLS

Abbreviation	Definition
AEOL	Aircraft Engine Operating Limits
AF	Air Force
AFB	Air Force Base
AFFTC	Air Force Flight Test Center
AFI	Air Force Instruction
AFOTEC	Air Force Operational Test and Evaluation Center
AFTO	Air Force technical order
AGL	above ground level
AOA	angle of attack
AOL	Aircraft Operating Limits
App	appendix
ASE	aero-servo elastic
CAS	Control Augmentation System
CDRLs	Contract Data Requirement List
CG	center of gravity
COMM/NAV	communication/navigation
CRL	Configurable Rail Launcher
CTF	Combined Test Force
DASD	Deputy Assistant Secretary of Defense
dB	decibel
DLL	design limit load
DOD	Department of Defense
DR	deficiency report
DT	developmental testing
DTIC	Defense Technical Information Center
DT&E	development, testing, and evaluation
EAFB	Edwards Air Force Base
EAR	Export Administration Regulations
EMD	engineering and manufacturing development
EMI	electro-magnetic interference
Encl	enclosure
etc.	and so forth
i.e.	that is
FFRR	First Flight Readiness Review
FLTS	flight test squadron
FOLD	Flight Operations Limitation Document
FSD	full scale development
FTCC	Flight Test Continuation Criteria

Abbreviation	Definition
G	gravity
GPS	global position system
GVT	ground vibration testing
HOTAS	Hands on Throttle and Stick
Hz	hertz
IADS	Interactive Analysis and Display System
IAW	in accordance with
ICD	Interface Control Document
INS	inertial navigation system
ITAR	International Traffic in Arms Regulations
ITB	Integrated Test Block
ITP	Integrated Product Team
ITT	Integrated Test Team
KPP	Key Performance Parameters
KSA	Key System Attributes
LCO	limit cycle oscillations
LDTO	Lead Development Test Organization
LEITB	leading edge integrated test block
LO	low observables
MIL-STD	military standard
MOA	Memorandum of Agreement
MUX	multiplex
NASA	National Aeronautics and Space Administration
NISPOM	National Industrial Security Program Operating Manual
NTIS	National Technical Information Service
OFP	operational flight program
OL	Operating Limits
OSD	Office of the Secretary of Defense
OT	operational testing
Pa	power approach
PAO	polyaphaolefin
PIO	pilot induced oscillations
PRV	pressure relief valve
P_S	negative specific excess power
psf	pounds per square foot
RCS	radar cross section
RTB	return to base
SAS	Stability Augmentation System
SDD	system design and develop
Sec	section

Abbreviation	Definition
SIL	System Integration Laboratory
SIM CERT	simulation certification
SMI	structural mode interaction
SOW	statement of work
TEMP	temporary
TENG	Test Engineering Group
TIH	technical information handbook
T.O.	technical order
TM	telemetry
TR	technical report
TSPI	time-space-position information
TRPR	Total System Performance Responsibility
TW	Test Wing
UAV	unmanned aerial vehicle
U.S.	United States
USAF	United States Air Force
U.S.C.	United States Code
Vol.	volume
V&V	Verification & Validation
WBS	work breakdown structure
WIT	watch item

This page was intentionally left blank.

APPENDIX F – DISTRIBUTION LIST (DRAFT)

Onsite	Number of Copies E-mail	Digital	Paper
Edwards AFB Technical Research Library Attn: Darrell Shiplett 307 E Popson Ave Edwards AFB CA 93524	0	0	2
AFTC/HO Email: jeannine.geiger@us.af.mil 305 E Popson Ave Edwards AFB CA 93524	1	0	0
773 TS/CL E-mail: john.kok@us.af.mil 307 E Popson Ave Edwards AFB CA 93524	1	0	0
412 TENG/CL E-mail: daniel.osburn@us.af.mil 307 E Popson Ave Edwards AFB CA 93524	1	0	0
412 OG/DT E-mail: christopher.klug.1@us.af.mil 195 E Popson Ave Edwards AFB CA 93524	1	0	0
412 TW/CT E-mail:412 TW Distribution Box 195 E Popson Ave Edwards AFB CA 93524	1	0	0
AFTC/CZ Attn: Joseph W. Nichols, PhD, SL 1 S Rosamond Blvd Edwards AFB CA 93524	0	0	1
Offsite			
Defense Technical Information Center Submit per DTIC procedures Attn: DTIC-O 8725 John J. Kingman Rd, Ste 0944 Ft Belvoir VA 22060 Email: aq@dtic.mil	1	0	0
Total	6	0	3

This page was intentionally left blank.