# INSUFFICIENTLY ROBUST DT&E MEANS TROUBLES AHEAD FOR OT&E

## *Unrealistic Operational Requirements Found*

*Raymond W. Reig*


In the fall of 1993, the General Accounting Office issued a report critical of how the Department of Defense (DoD) was managing the acquisition and development of software-intensive systems.[1] At about the same time, the Under Secretary of Defense (Acquisition and Technology) (USD(A&T)) asked his staff how come systems pass developmental testing (DT), and fail operational testing (OT). He then named five systems that were electronic warfare (EW) and command, control, communication, computer, and Intelligence (C$^4$I) systems. These systems are, of course, software intensive; thus, the actions are addressing the same issue.

An intensive three- to four-month study effort was initiated to answer the USD(A&T) question involving several organizations and considerable number of personnel in the test and other acquisition disciplines. The Defense Systems Management College Test and Evaluation Department participated in research aimed at answering the question. Conclusions were presented through management and considered with many other inputs into the final report. This report was forwarded to the USD(A&T) on 25 February 1994, and contained the following primary findings:[2]

— The requirements generation and management process led to unrealistic operational requirements.

— Program Developmental Test and Evaluation (DT&E) was not sufficiently robust for confident entrance into Operational Test and Evaluation (OT&E).

— System boundaries were not defined sufficiently.

My involvement with this study effort, combined with other experiences in weapon systems acquisition, have resulted in "personal findings," as follows:

*Professor Reig is on the faculty of the Test and Evaluation Department at the Defense Systems Management College. His more than 35 years of experience spans military, government and private aerospace industry.*

### *Personal Findings*

1. The requirements generation and justification documents — the Operational Requirements Document (ORD) and the Cost and Operational Effectiveness Analysis (COEA) — tend to be used as advocacy documents leading to optimistic expectations, which are rarely achieved. Research into the acquisition history of 24 DoD programs shows an average cost overrun in the Engineering and Manufac-

turing Development (EMD) phase of the programs to be 45 percent, and the schedule overrun to be 63 percent.[3] A natural bias seems to be at work here. After all, no one proposes a new system "just a little better" than the existing system. So the pressure builds for the super system that will stand head and shoulders above the existing system. This usually requires state-of-the-art solutions and results in notching up the risk. That could take the form of adding one more requirement to the existing hard-core


new requirements, thus leading to the one super system that can accomplish the entire requirement. In simpler times, this was called goldplating and engineers were accused of doing it.

Concurrently, the COEA is accomplished and "confirms" that to meet all requirements the system in mind is most cost-effective. The COEA that comes to mind included a major Pre-Planned Product Improvement ($P^3I$) consisting at the time of a one- or two-sentence description of the modification, and the specific threat it would defeat. Regardless of the fragile input data, the COEA concluded that the basic system with the $P^3I$ was the preferred approach. Independent cost analysis concluded insufficient data existed to determine the reasonableness of funding profiles of the basic program, to say nothing about the $P^3I$ program.

Another example of how the ORD and COEA complement each other is the V-22, Tilt-Rotor program. The original COEA used an operational scenario representative of past and current tactics.

This resulted in an uncomfortable level of cost. Now, my understanding is that a new COEA is being created, using the newer and future operational concept of Operational Maneuver from the Sea. This should increase the cost-effectiveness of the system and may correct a prior fault, but it shows also how the ORD and COEA tend to self-confirm each other. Lastly, the DoD 5000 series acquisition guidance makes the user responsible for creating both the ORD and the COEA. This single responsibility for the requirement, as well as the justification of the selected solution, is rare in the practice of business.

2. Developing successful software-intensive systems is difficult, because more hardware experience and skills are available in industry and government than software experience and skill. This is particularly true in older, senior personnel who started their careers in the 1950s and 1960s — the "hardware generations." Back then, in large engineering departments working on many programs, one centralized compartmented area existed, where we engineers would bring software requirements to the counter, and were told when to return for the software program to integrate into the hardware at a final manufacturing step. When we returned, we usually were told the program wasn't ready, when to return again, and how much more funding to bring. (An over-simplification, but not by much.)

**DT is considered a "low hurdle" (easy), while OT is considered a "high hurdle" (harder). This is especially true for software-intensive systems.**

Much progress has been made in software development since then, much of it codified in MIL-STDs

2167A and 2168. But, hardware engineers who used to bring their requirements to the counter have not received much additional training. Many of those still in the business are faced with making decisions on software-intensive systems. This results in managers not skilled in the subject area deciding major program design parameters. However, on balance, they come with a healthy skepticism of software plans, based on their frequent return trips to the software counter.

3. Avionics software-intensive systems are even more difficult to develop successfully, because airborne equipment usually has severe state-of-the-art requirements involving weight, volume and cooling air. One manifestation of this is low volume manufacturing yields of densely packaged electronics. The same card, made in the engineering lab for proof of principle tests on the prototype, did not reveal this manufacturing problem. This led to optimistic production schedules.

Also, unhampered by fact, it is my opinion that avionics software-intensive systems have more interfaces than stabile ground systems. Shortcomings are more apt to be discovered in OT end-to-end testing than in prior *in situ* DT testing. For example, a standard radio designed to operate in many aircraft types may perform differently in different platforms because of the type and placement of the antennae in the various platforms, particularly in fighter aircraft vs. transport aircraft.

4. The EW and C⁴I avionics software-intensive systems are the most difficult to develop successfully. In addition to the constraints previously discussed, these systems usually require stringent special personnel security clearances, over and above those required for other military development projects.

In the late 1960s and early 1970s, the "software generations" were just graduating with the prerequisite skills desperately needed but, to a certain percentage, using their skills for the military was anathema. Any involved in college student protests probably would not be granted the required clearance. Besides, the alternative was employment in Silicon Valley and elsewhere with modern working conditions and attractive pay. It seemed more inviting than working at a green or gray steel desk in an engineering bullpen at some defense contractor or government laboratory.

Also, the EW and C⁴I systems were being designed to defeat a threat that was not static, not under our control, or perhaps not even known to us. In the 8-10 years it takes to design, develop and produce our system, if the threat has increased, our production-ready system may have limitations. This comes close to being a law of physics and the only solution I can think of is a P³I or evolutionary acquisition approach.

5. Perhaps because of the aforementioned aspects, software-intensive systems receive extramanagerial inputs, above and beyond the program manager's control. Examples listed here are all from one program, but these and others do occur with some frequency. In this program, the Service Secretary abruptly requires unplanned-for testing, and adjusts the budget insufficiently. Another Service withdraws from this joint program, and Congress limits the use of production funds. Contracting directs the procurement of a critical subsystem from a specific source, and later caps the government program costs at a figure substantially below actual costs. Most recently, the Service has been directed to conduct no more OT until a Director of Operational Test and Evaluation, has been appointed.

6. In systems of this type a movement seems to exist toward "creative terminology" which confounds the established program evaluation metrics. I believe most people who have been in the defense industry for any period of time have a good, homogeneous interpretation of the terms Low Rate Initial Production (LRIP), Operation Evaluation (OPEVAL), Full Rate Production, and Initial Operational Capability (IOC). Generally, they are familiar with the requirements of each, and many of these requirements are contained within the DoD 5000 series documentation. Recently, some programs approaching these milestones, but not quite able to meet the requirements thereof, have used alternative terminology. Hence, an LRIP phase becomes a PV (product verification) phase, or an OPEVAL becomes an "Operational Effectiveness Test" or a "Verification of Correction of Deficiencies Test." The IOC, in some cases, has been replaced by Limited Operational Capability (LOC). Many people would know how to evaluate an OPEVAL, or judge readiness for an LRIP phase, but no guidance is found on how to evaluate, or what to expect from, these newer terms.

Then the question remains regarding wording used within a TEMP or test report. One sentence from a test result used in a program TEMP stated "For the tests performed the system operated as required." Is this good or bad? Another variation of the same idea is to list a large number of limitations of scope on the testing performed, and then provide a generalized evaluation which makes the results of limited usefulness. Currently a program is considering declaring IOC before Milestone III, upon delivery of the first LRIP article. Clearly, this is a different interpretation of the IOC than that of a few years ago.

7. In my opinion, subject to objective confirmation, it seems that few test articles have been used in the EMD phase, relative to the total planned production quantity. In the ongoing research into the EMD phase of recent weapon systems acquisitions, an average of 1.8 percent of the total planned production quantity or

28 percent of the total LRIP quantity was acquired with research, development, test and evaluation funds and presumably used for testing. Of the five programs used as examples in USD(A&T) questions, the average number of LRIP articles used for testing was 1.3 percent of the total planned production quantity.[4]

**Recommendations**

The price for commenting on a process is to recommend actions to ameliorate the situation. Mine are as follows:

1. If someone could solve the software development problem, half of our systems acquisition problems would be solved. Our usual approach is to reorganize or change the acquisition policy. I suggest a complementary approach whereby, for the next five years, a mandatory special software acquisition management course could be held for all senior managers involved in DoD acquisition. The target audience would be the still-active hardware-generation managers. Waivers would be available for the software proficient, and younger managers who want more software training would be welcome. After five years the problem should be self-correcting, and the special course could be terminated.

The Air Force Bold Stroke course is an excellent example. This is a one-week course designed for general officers, and its objective is to increase the awareness of these senior managers to software acquisition pitfalls. This course, originally designed by DSMC for the Air Force Systems Command, is now presented by the Professional Development Institute.

2. Distinguish between a "broken" acquisition system and a poorly executed program. In the 24 systems reviewed, 18 were tightly grouped somewhat over the original EMD cost and schedule estimates, and six were significantly outside the range. I would say the current acquisition system provides results similar to the results of 18 of these programs, and six of these programs were poorly executed. The reasons for poor execution are many and diverse, and probably caused as much by conditions outside the program office as those within.

3. Instill discipline in the established acquisition system. The current official guidance governing DoD acquisition, the 5000 series documentation, was issued in February 1991. Merely reading and knowing the 5000 series intimately would not suffice three years later. The conclusions of a dozen or more OSD memos, immediately directive in nature, indicate that the subject matter will be incorporated into the next Directives update. To date, this has not happened.

The Directives state the Test and Evaluation Master Plan is limited to 30 pages, plus annexes. Providing the data currently required by OSD reviewers and others within 30 pages is not possible. Much of the additional required data is valuable, and perhaps the answer, in this instance, is to review the page limitation. Creative terminology, discussed previously, falls into this category of system discipline.

4. Consider the then-current practices when *post-facto* criticizing a program's performance. I believe the A-12 Administrative Inquiry contains an eloquent exposition of this point.[5]

The above thoughts were generated by my involvement in recent research efforts associated with acquisition, and by longer-standing experience. I hope it agrees generally with your own thoughts, and you find it worthwhile. The Test and Evaluation Department research is ongoing and other resulting data will be published in subsequent articles.

**Endnotes**

1. "Test and Evaluation: DoD Has Been Slow in Improving Testing of Software Intensive Systems," 28 September 1993. (GAO/NSIAD-93-198), Washington, D.C., General Accounting Office.

2. Wiles, J. A. (25 February 1994). "Study Group Report on Evaluation of Electronic System Acquisition," Report to the Under Secretary of Defense (Acquisition and Technology). Washington, D.C.: Department of Defense.

**The Air Force Bold Stroke course is a one-week course designed for general officers, and its objective is to increase the awareness of these senior managers to software acquisition pitfalls.**

3. Gailey, C. K., R. W. Reig, and W. Weber (1994). Ongoing, unpublished, DOTE-sponsored research of DoD acquisition. Test and Evaluation Department, Defense Systems Management College, Fort Belvoir, Va.

4. *Ibid.*

5. Beach, C. P., Jr. (28 November 1990). " A-12 Administrative Inquiry." (Report to the Secretary of the Navy) Washington, D.C.