ITEA Journal 2013; 34: 158–163

Cluster Testing: Reducing Air Weapon Test Execution Cycle Time

George Axiotis
Department of Defense, Arlington, Virginia
Steve Riker
Foxhole Technology Corporation, Fairfax, Virginia

In an era of rapidly changing threats, the Department of Defense (DoD) strives to find more effective ways to rapidly field capability while ensuring that new capabilities are adequately tested prior to acquisition. This article proposes a departure for air weapon test programs from the traditional serial shoot–analyze–shoot approach with slowly increasing complexity to a clustered test phase approach of compressing test points of varying complexity as a way to reduce overall test timelines without significantly losing test point knowledge.

Key words: Test & evaluation; DT; OT; developmental test; weapon testing; test centers; rapid fielding; test cycle time; design of experiments; test points; Axiotis; Riker.

The Department of Defense (DoD) currently emphasizes rapid fielding of incremental capability improvements to counter changes in the threat. Although software-based weapon systems have done much to facilitate staying ahead of the threat, test program duration remains a pacing item to fielding new capability. This article examines current air-to-air (ATA) and air-to-ground (ATG) live weapon test methodologies that must now adapt to meet truncated test programs and urgent fielding needs. The authors propose a more streamlined approach by departing from today's accepted shoot-analyze-shoot processes with incremental complexity to a more efficient clustering of test events of varying complexity at a more economical rate. The authors found that clustering combined with randomized test events not only leads to faster and more effective discovery and shortens development timelines through faster test cycles but more importantly it does so without significant loss in test point knowledge. The value of this process is in addressing program risk early, allowing sufficient time for resolution of issues prior to operational test (OT).

SR CASE 13-S-0466

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Director, Developmental Test and Evaluation, the Department of Defense, or the U.S. Government.

State of weapon development test

Although weapon complexity has substantially increased in the last 20 years, the DoD conducts fewer live air weapon tests (defined as full-up dynamic flights with or without warheads) today than it did 20 years ago. A RAND Corporation study of nine key ATA and ATG weapon programs from 1980 to 2000 showed an average of 15 months of system development followed by 60 months of combined developmental test (DT) and OT with an average of 33 DT live shots for ATA and 43 DT live shots for ATG weapons expended (Fox et al. 2004).

Today, the combined average of the eight main DoD ATA and ATG weapon programs' DT and integrated test (IT) periods is a bit shorter at 48 months for new starts and roughly 24 to 30 months for major upgrades with a typical 6-month period for OT. The number of flight test assets, however, has reduced to an average of 19 live shots in DT with 20 for OT (Axiotis et al. 2011).

The reasons for these reductions are partly due to advances in modeling and simulation but also because the cost of live-fire weapons has tipped the balance more to captive carry and simulations. In fixed-budget environments, increased flight test weapon costs force managers to procure fewer assets for testing, and each shot becomes that much more important in the final assessment. There is a drive to maximize the use of limited test assets by conducting IT that allows data to be used in both DT and OT evaluations. As testing

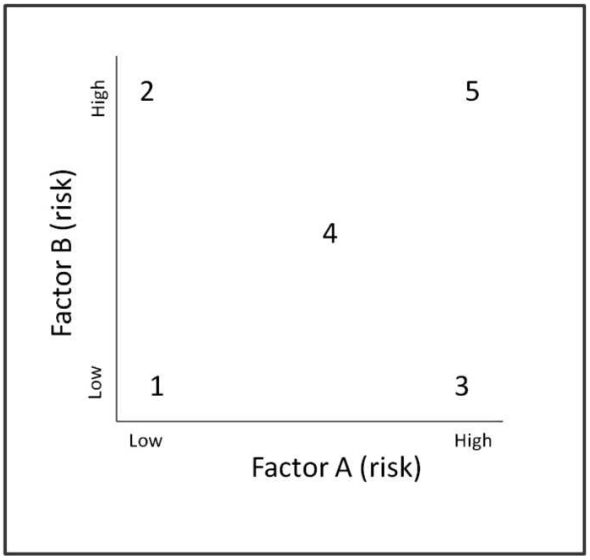

Figure 1. Traditional test space construct.

shifts in favor of IT, we expect risk to be advanced in order to meet the criteria demanded in OT. Additionally, in response to limited and costly test asset allocation, the DoD now emphasizes the use of analytical test planning tools such as design of experiments to determine the minimum number of tests necessary to confidently and efficiently assess the interactions of performance-influencing factors. A similar approach to test execution is also needed if we are to further shorten DT periods.

Current test execution methodology

Current test and evaluation (T&E) planning methodology and practice for weapon testing involves executing a mix of "dummy round" tests to examine weapon separation and ballistic effects and captive-carry tests to assess sensor and some logic performance to validate design models and then augmenting those tests with live tests to include full-up instrumented and warshot rounds. The latter group is specifically structured to assess full end-to-end (i.e., kill chain) performance, focusing on launch and flight dynamics, weapon-sensor interactions, endgame logic, and fusing. This article focuses on the live flight test periods. Although the analytical methods described previously have done much for test planning, we still execute the tests themselves in a methodical, but slow, shoot-look-shoot serial process. The underlying methodology for test execution is to conduct a serial test series beginning with relatively low-risk test points in the heart of the envelope and then progressing with ever-increasing complexity to reach key performance parameters. Weapons that are designed with multiple operational modes also follow this methodology, resolving issues in the less complex modes before attempting to resolve those in the more complex modes, further delaying the detection of critical performance issues. *Figure 1* represents a notional test space in which a multimode weapon system is being tested in a traditional construct. The low-risk approach would imply that basic functions are tested early (in the sequence depicted) with risk added sequentially as performance is demonstrated.

This methodical process allows risk to be managed by limiting the influencing factors during test as well as by testing functionality as it matures. Beginning slow and expanding complexity over time, though risk averse, unfortunately pushes key discovery to late in the test program where there is little time for rework and retest. The periods between test shots, called test centers, comprise the analysis periods from prior shots. Range, weapon, instrumentation, and target preparations; certifications; and manning are necessary to execute subsequent testing. Historically, weapon tests averaged between 0.5 to 0.7 shots per month (6- to 8-week test centers) depending upon weapon complexity. Today, for the marquee ATA and ATG weapon programs within the DoD, the test planning factors remain nearly the same at 0.5 to 0.9 shots per month for ATG weapons (depending on whether the weapon is powered) and 0.5 shots per month for ATA weapons (DASD-DT&E Air Warfare Portfolio Review 2011). For a typical 36- to 48-month weapon development process, the live test program for a 10-shot series would require an additional 14 to 20 months before considering entering OT. In a rapidly changing threat environment, a 2-year program extension to complete DT and OT can have significant consequences.

Reducing test cycle time

Reducing overall DT shot cycle time, given a constant knowledge requirement, would require either accelerating preparatory activities (shortening test centers), reducing the total number of test shots by significantly increasing the complexity of each event, increasing the number of shots conducted per test period, testing separate operational modes concurrently, or some combination of these options.

Accelerating preparatory activities to achieve the kinds of savings envisioned introduces higher risk primarily due to human error and in many cases is impractical due to fixed cycle time for other related assets such as targets and instrumentation. In the second case, complexity analysis and experience by testers indicates that as test complexity increases, the less likely its probability of successful execution (Axiotis 1995). The desire for statistically confident data sets makes the first two alternatives risky. Given

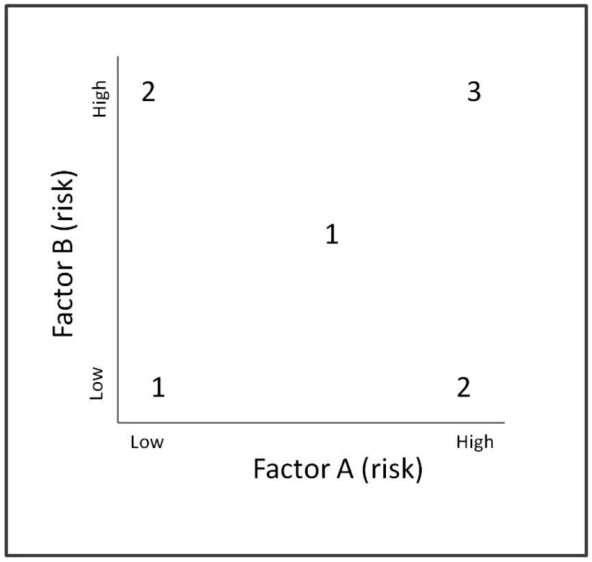

Figure 2. Cluster test space construct.

the cost of field and range tests, testers opt to design manageable scenarios and hold to conservative test asset preparation times to ensure reasonable test success. That leaves rethinking how many test shots could or should be conducted per test period as a way to possibly reduce overall development and test cycle times.

Proposal

The authors propose two key process evolutions from the established shoot-analyze-shoot approach evolving from simple heart-of-the-envelope events out to more rigorous edge-of-the-envelope tests. The first evolution increases the number of test shots in each test period, and the second requires randomizing the test event complexity within a mature functional build consistent with design of experiments precepts of factor interactions. The objective is to create more efficient test "clusters," providing more data to support design evolution during the subsequent test center periods. This clustering mimics that used in traditional sampling techniques where a population is partitioned into subgroups for test. In traditional design of experiments, this can be also seen as a nested test design. Traditional cluster sampling produces estimates with large variances (Feinberg 2003), but in this proposal, the objective is to use test clusters of a limited configuration range to build a larger pool of data per test event. These test clusters would start small and increase over the test program with corresponding increases in test period duration until the test matrix is complete. Test complexity is also randomized throughout the execution plan, as objectified in a design of experiments framework, to look for key interactions earlier in the test cycle. The key aspects of the clustered test approach are the following:

- design of experiments methods build the overall test matrix;
- live missile test events are grouped with complexity suitably randomized into clusters;
- initial test clusters are relatively small exercised with a mix of scenario complexity; and
- subsequent test cluster size increases with mixed scenario complexity.

The objectives of this approach are threefold: (1) stress the system earlier and more often, (2) gather more data per test cycle across a greater range of a set of factors to support more meaningful analysis between test centers, and (3) limit the impact of a single failure during test.

Figure 2 represents a notional test space in which a multimode weapons system is being assessed. Here, risk is introduced earlier in a cluster with lower risk test points. The goal is to drive discovery in test earlier in the test sequence to allow fixes (if required) to be more efficiently introduced and proven in subsequent tests. To demonstrate whether a "clustered" test series is inherently more efficient, a notional example is assessed.

Test shot efficiency: notional example

Tables 1–3 compare test duration and test point knowledge achievement between a notional typical serial test series and two clustered test options.

Assumptions: 6-week centers, 2 days per test event, refly rate decrease over time. One hundred knowledge point objective:

Table 1. Standard: shoot-analyze-shoot (with progressive difficulty) (one shot per 6-week center).

	Shot 1	Shot 2	Shot 3	Shot 4	Shot 5	Shot 6	Shot 7	Shot 8	Shot 9	Shot 10
Weeks	**0**	**6.4**	**13**	**19.4**	**26**	**32.4**	**39**	**45.4**	**52**	**58.4**
Test complexity	1	1	1	2	2	2	2	3	3	3
Test points planned	5	5	5	10	10	10	10	15	15	15
Refly rate	40%	40%	40%	30%	30%	30%	25%	20%	20%	20%
Test points Achieved	3	3	3	7	7	7	8	12	12	12
Total test points	**3**	**6**	**9**	**16**	**23**	**30**	**38**	**50**	**62**	**74**

Table 2. Cluster option 1 (1×2-shot + 2×4-shot periods).

	Shot 1	Shot 2	Shot 3	Shot 4	Shot 5	Shot 6	Shot 7	Shot 8	Shot 9	Shot 10
Cluster	**1**		**2**				**3**			
Weeks		**8.4**				**21**				**33.6**
Test complexity	1	2	1	3	3	2	1	2	3	2
Test points planned	5	10	5	15	15	10	5	10	15	10
Refly rate	40%	40%	40%	30%	30%	30%	25%	20%	20%	20%
Test points achieved	3	6	3	10	10	7	4	8	12	8
Total test points	**3**	**9**	**12**	**22**	**32**	**39**	**43**	**51**	**63**	**71**

- total test period: 58.4 weeks;
- at roughly 50% through the test program (week 26), achieved less than 25% (23/100) of the knowledge point objective;
- at 75% through the test program (week 45), achieved 50% of the knowledge point objective and left the significant unknowns in the most complex events.

Assumptions: 8-week centers for two-shot events, 12-week centers for four-shot events; 2 days per two- or three-shot test event, 3 days per four-shot event; refly rate decrease over time. One hundred knowledge point objective:

- total test period: 33.6 weeks;
- at 50% through the test program (shot five), achieved 32% of the knowledge point objective;
- by shot 10 (week 33.6), roughly the same amount (<5% difference) of the total knowledge point objective is achieved as in the standard test series; complex events exercised earlier.

Assumptions: 8-week centers for three-shot events, 12-week centers for four-shot events; 2 days per three-shot test event, 3 days per four-shot event; refly rate decrease over time. One hundred knowledge point objective:

- total test period: 30 weeks;
- at 50% through the test program (shot five), achieved 28% of the knowledge point objective;
- by shot 10 (week 30), roughly the same amount of total knowledge point objective is achieved as in option 1; complex events exercised earlier.

Observations

In the notional examples, as expected, both clustered test options showed that significant test schedule reductions were possible using this approach compared with the traditional shoot-analyze-shoot method. This reduction was to be expected given the overall strategy of increasing shots per test phase. More importantly, the examples show the clustered test options gaining test point knowledge at a faster rate with no significant loss of overall test point knowledge by test completion. Finally, an additional assumed benefit will be the inherent increased time between test centers, allowing for more complete analysis of test data and robust corrective action periods. These findings must now be examined using a real-world acquisition program example.

Program example

The next step is to apply both key tenets of this article to a program example. In this case, the Weapon X program (actual system description and data altered) test matrix, as shown in *Table 4*, reflects 10 live shots exercised over 15 months based on one shot every 6 weeks. The test difficulty (test rating) is reflected in the far right column. The test matrix uses a design of experiments framework based on a full-factorial design with suitable power, with an approximate 80% confidence level. The test matrix screens for important factors driving performance across multiple conditions. Based on interactions, the matrix is efficient from a design of experiments perspective but not necessarily for efficient test execution. The program's test matrix is executed using a traditional crawl-walk-run method-

Table 3. Cluster option 2 (2×3-shot + 1×4-shot periods).

	Shot 1	Shot 2	Shot 3	Shot 4	Shot 5	Shot 6	Shot 7	Shot 8	Shot 9	Shot 10
Cluster	**1**			**2**			**3**			
Weeks			**8.4**			**17**				**30**
Test complexity	1	2	3	1	2	3	1	2	3	2
Test points planned	5	10	15	5	10	15	5	10	15	10
Refly rate	40%	40%	40%	30%	30%	30%	25%	20%	20%	20%
Test points achieved	3	6	9	3	7	10	4	8	12	8
Total test points	**3**	**9**	**18**	**21**	**28**	**38**	**42**	**50**	**62**	**70**

Table 4. Weapon program example shot matrix (based on design of experiments).

Shot	Look mode	Lock mode	Range	Quarter	Cue type	Test difficulty		Test rating
						Launch (1:5)	Target (1:5)	
1	A	A	L	F	A	1	1	2
2	B	B	L	R	C	1	2	3
3	C	B	L	B	C	3	2	5
4	C	B	L	F	B	3	2	5
5	B	C	S	B	A	4	1	5
6	A	B	S	F	B	2	3	5
7	C	B	L	F	C	3	3	6
8	B	A	S	B	A	3	4	7
9	C	A	S	B	A	4	4	8
10	C	B	S	F	A	5	4	9

ology of increasing difficulty with one shot every 6 weeks over a period of 60 weeks.

We apply the same clustering approach to assess whether the planned 15-month test program could be reduced while maintaining roughly the same overall knowledge. As shown in *Table 5*, without clustering, the matrix reflects the knowledge gained at test midpoint at less than 25% of the objective. *Table 6* shows that with clustering, the same knowledge could be achieved in less time, even with shot randomization.

Assumptions: 6-week centers, refly rate decrease linearly over time. One hundred knowledge point objective.

Assumptions: 12-week centers for four-shot events; 2 days per two-shot test event, 3 days per four-shot event; refly rate decrease over time, first shot requires more refly. One hundred knowledge point objective.

Findings

Based on the program example, clustering tests with shot randomization has the potential for significant test efficiency. Even with two successive shot randomizations (not shown), overall knowledge achieved was within 10% of that for the traditional method. Knowledge achieved was mathematically shown, based primarily on the assumptions used for knowledge efficiency. For practical purposes, this approach has its limits as it assumes system functionality is close to final at time of test and relatively constant over time. This assumption may not be the case as initial software releases generally exercise only parts of overall expected functionality. Thus, for some programs, pushing complex test events earlier in the test matrix may not be possible or even desirable. This construct will have to be verified through actual program test matrix execution. At this time, one program is considering such a test clustering approach.

Conclusions

Through rethinking how we gather knowledge, we have the potential to accelerate weapon test programs through test clustering. These clusters decrease the total number of test periods by conducting multiple shots per test phase and pushing discovery earlier by randomizing test event complexity within a configuration cluster. This approach can be accomplished without significant loss of overall knowledge. Introducing a cluster test strategy will require a culture change at the program manager level. As risk is advanced in the test process, program managers may have to defend their programs earlier in development when tests reveal deficiencies, even though they may now have more time to address 'fixes." In an austere fiscal environment, there will be a natural tendency to reduce risk so as not to be vulnerable to adverse budget action from test "failures." Accepting increased risk

Table 5. Program example—no clustering/no randomization.

	Shot 1	Shot 2	Shot 3	Shot 4	Shot 5	Shot 6	Shot 7	Shot 8	Shot 9	Shot 10
Weeks	**0**	**6.4**	**13**	**19.4**	**26**	**32.4**	**39**	**45.4**	**52**	**58.4**
Test complexity	2	3	5	5	5	5	6	7	8	9
Knowledge points Planned	6	7	8	8	8	8	12	13	15	15
Refly rate	40%	40%	40%	30%	30%	30%	25%	20%	20%	20%
Knowledge points achieved	3	4	5	5	5	5	9	10	12	12
Total knowledge points	**3**	**7**	**12**	**17**	**22**	**27**	**36**	**46**	**58**	**70**

Table 6. Program example—cluster option 1 with shot randomization.

	Shot 1	Shot 2	Shot 3	Shot 4	Shot 5	Shot 6	Shot 7	Shot 8	Shot 9	Shot 10
Cluster	**1**		**2**				**3**			
Weeks		**8.4**				**21**				**33.6**
Test complexity	5	9	5	8	5	2	7	6	3	5
Knowledge points Planned	8	15	8	15	8	6	13	12	7	8
Refly rate	40%	40%	40%	30%	30%	30%	25%	25%	20%	20%
Knowledge points achieved	5	9	5	10	6	4	10	9	6	6
Total knowledge points	**5**	**14**	**19**	**29**	**35**	**39**	**49**	**58**	**64**	**70**

may be more efficient but runs counter to the program management instincts. ❑

George Axiotis is the deputy director for Air Warfare Systems within the Office of the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation. His portfolio includes more than 60 tactical and support aircraft, sensors, electronic warfare systems, as well weapons. He formerly led test resources analysis within the Office of the Director of Operational Test and Evaluation and prior to that served as a director of a Navy Material Command T&E Office. Mr. Axiotis has more than 29 years in defense acquisition and T&E having been a project and test manager on Defense programs, T&E advisor to two PEOs, as well as a program resource sponsor. E-mail: george.axiotis@osd.mil

Steve Riker is the lead ATA/ATG weapons and Navy tactical air analyst within the Office of the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation. His previous assignments included test resource manager at the Missile Defense Agency and Operational Test Director for the Navy F-4 series aircraft at Air Test and Evaluation Squadron Four (VX-4). E-mail: steve.riker@osd.mil

References

Axiotis, G. "T&E Roulette, Where Is the Smart Bet?" Presentation to the Association of Scientists and Engineers, Naval Sea Systems Command, 1995.

Axiotis, G., and Steve Riker. "Air Warfare Systems Portfolio Review," DoD DASD-DT&E, July 2011.

Feinberg, Stephan. *Cluster Sampling*. Department of Statistics, Carnegie Mellon University, Pittsburg, PA. www.stat.cmu.edu/~fienberg/Stat36-303-03/clustering-rev03.pdf. 2003. Accessed October 1, 2012.

Fox, B., M. Boito, J. Grasser, and O. Younossi. 2004. "Test and Evaluation Trends and Costs for Aircraft and Guided Weapons," RAND Corporation Study, 2004.

Summary brief to the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation, July 2011, DASD-DT&E Air Warfare Portfolio Review.