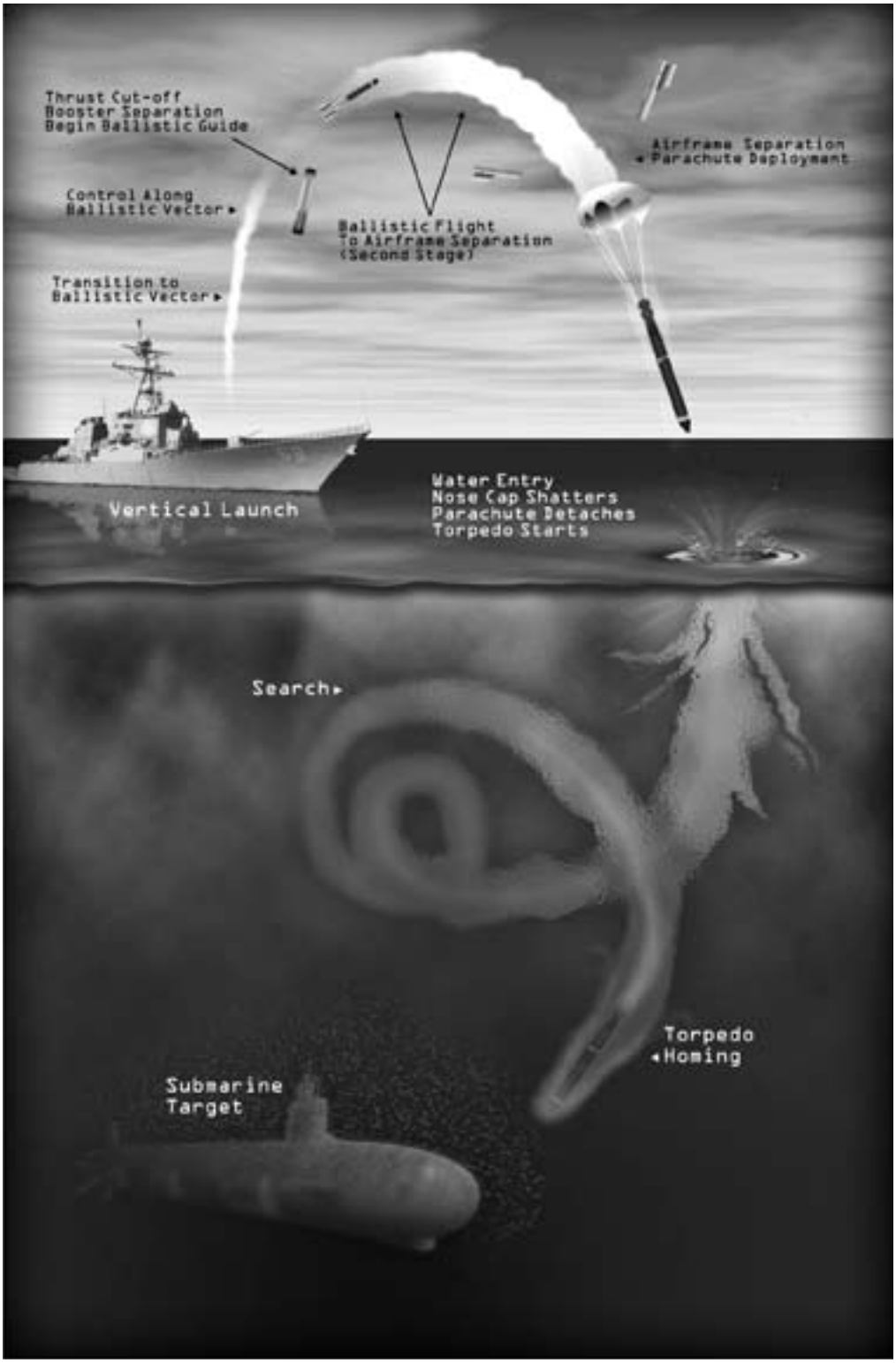

Image designed by TSgt James Smith, USAF

TEST AND EVALUATION LESSONS LEARNED FROM THE FIELD

Karen M. Stadler

This article examines test and evaluation (T&E) lessons learned from more than 300 students with extensive T&E field experience who attended the Defense Acquisition University (DAU) test and evaluation classes during FY02–FY05. The T&E lessons learned in 18 categories were researched and correlated, and findings in the top five categories are presented. In particular, this article focuses on detailed lessons learned in the areas of test design and execution, test planning, teamwork and communication, funding, and scheduling. A compilation of student (field practitioner) comments and recommendations is presented, and overall results are compared with results from other similar studies and documents.

As part of the DAU Advanced Test and Evaluation (TST 301) class, students prepare and present PowerPoint slideshows on various T&E-related topics. Many students prepare and present detailed briefings on their T&E lessons learned. The lessons-learned presentations are typically based on actual experiences in planning, conducting, analyzing, and reporting test results. Students typically have many years of T&E/acquisition field experience, and their presentations contain a wealth of valuable information, which could help others avoid common sources of error when designing and executing test events. The purpose of this article is to identify and discuss common T&E best practices and lessons learned, thereby enabling possible cost and schedule savings and improved test results from future T&E efforts. Decision makers and acquisition/program leadership can benefit from this article by better understanding the top T&E related issues, as reported by field practitioners.

The T&E lessons-learned data was obtained from students who attended this author's TST 301 classes during FY02-FY05. Of the 393 students, 301 students

TABLE 1.
NUMBER OF STUDENTS BY ORGANIZATION/SERVICE

ORGANIZATION/SERVICE	NUMBER OF STUDENTS	PERCENT OF STUDENTS
U.S. Army	141	46.8
U.S. Air Force	89	29.6
U.S. Navy	35	11.6
Missile Defense Agency (MDA)	24	8.0
Special Operations	6	2.0
Department of Defense (DoD)	4	1.3
U.S. Marine Corps	2	0.7
Industry/Contractor	0	0.0
Total Number of Students	**301**	**100.0**

presented significant T&E lessons-learned information. The students came from all four services and DoD agencies, as summarized in Table 1. The lessons-learned data consists solely of student comments and opinions and is based on student knowledge and experience in the T&E area, along with any research conducted by individual students.

The student data was analyzed for common trends, and 18 different categories, covering all common trends, were selected. The student briefings were then tabulated, to determine the number of student briefings with lessons learned for each of the 18 categories. For example, 192 of the 301 total students (63.8% of the total) had significant lessons learned in the area of test design and test execution. Note that each student briefing contained lessons learned from one or more of the 18 categories. The student lessons learned data is presented in Table 2.

MAJOR FINDINGS

Of the 18 categories of T&E lessons learned in Table 2, this article will further discuss the top five categories (test design and execution, test planning, teamwork and communication, funding, and schedule). Note that far more students (41%–64%) had lessons learned in the top five categories than in the bottom 13 categories (5%–30% of students). This article contains a summary and detailed student comments and recommendations for each of the top five areas. This information can benefit the T&E community by providing detailed lessons learned, which might assist future T&E efforts and help acquisition leadership better understand the major T&E issues and concerns.

TABLE 2.
NUMBER OF STUDENTS WITH LESSONS LEARNED, IN EACH CATEGORY

LESSONS LEARNED CATEGORY	NUMBER OF STUDENTS	PERCENT OF STUDENTS
1. Test design, test methods, test execution, and analysis methods	192	63.8
2. Test planning	162	53.8
3. Teamwork and communication	141	46.8
4. Funding, budget, and cost	141	46.8
5. Schedule	122	40.5
6. Test infrastructure, test tools, test articles, and instrumentation	91	30.2
7. Test requirements	91	30.2
8. Safety and risk management	85	28.2
9. Government leadership and management issues and organizational politics	66	21.9
10. Contractor issues, including contractor leadership and management issues	55	18.3
11. Training issues	35	11.6
12. Modeling and simulation (used in conjunction with testing)	34	11.3
13. Interfaces, interoperability, and integration issues	33	11.0
14. "Stuff happens" (Murphy's law, weather)	32	10.6
15. Manpower issues	27	9.0
16. Immature technology and/or immature system	15	5.0
17. Commercial-off-the-shelf (COTS) and Non-development Item (NDI) issues	15	5.0
18. Poor judgment	15	5.0
Total Number of Students	**301**	**100.0**

THE TOP FIVE STUDENT CATEGORIES

TEST DESIGN, TEST METHODS, TEST EXECUTION, AND ANALYSIS METHODS (63.8 PERCENT)

Test design, methods, execution, and analysis methods are unquestionably among the largest factors that determine the success or failure of test events. Students offered the following advice in this area:

- Take the time to develop a robust T&E strategy and to determine the best (anticipated) test design and methods for your situation. Consult experts where necessary. Examine all facets of T&E such as instrumentation, data collection, analysis methods, test validity, test procedures, etc. For example, one might consider a side-by-side comparison of the existing and the new system. The new system may not meet all the requirements, but may be significantly better than the existing system. Without the comparison, the true conclusion may be missed.

- The devil is in the details. Student after student indicated that the little details (as well as the overall test design and execution) greatly affected the success or failure of their test events. For example, one detail of test planning might be to consider collecting diagnostic data, as well as test data. In case of equipment failure, the data can assist in determining the problem(s).

- Understand the test objectives, including how and why the test requirements were generated. Know the *what* and *why* behind limits and guidelines. Look at the system requirements, don't just accept them. Spend the time upfront with users and requirements developers to get the requirements well defined, especially where the requirements don't make sense. Often, the user may not know what he/she wants and why. Is the test relevant? Are the requirements realistic? The user will sometimes change the requirements, if the need is explained to them. Systems engineers and testers need to work together.

- Include tests at realistic operating conditions and at all corners of the envelope. These tests should occur prior to production or as early as possible. Measure all critical parameters and verify all requirements. Review the Test and Evaluation Management Plan (TEMP) and test plans versus requirements (e.g., was essential data collected and nonessential data not collected?).

TEST PLANNING (53.8 PERCENT)

Numerous students stressed the importance of thorough test planning. There are things one cannot or will not anticipate; stuff happens, people make mistakes. But proper anticipation and thorough planning will result in fewer problems in the long run, including a better chance of success and remaining within cost and on schedule. Good planning allows proper resource allocation and makes test execution far easier. Additional student comments concerning the importance of test planning and not cutting corners are as follows:

- Any test plan should have more than one person to review it. Make use of available expertise.

- Do analysis prior to testing to try to predict and anticipate results. This analysis will assist in identifying potential problems and developing contingency plans. It also helps identify needed changes to test plans.

- Plan for contingencies, especially weather. Have an alternate test plan/alternate test points available for each day of testing.

- When planning a test with others, start early. They were already busy before you came. Attempt to keep things simple. Questions should be clear and to the point. Follow their processes, if possible.

- Become an expert on systems you are testing. Tap into subject matter expert (SME) knowledge. Spend time with the user. Find out the user's priorities, concerns, and the reasons for these. Consider traveling to the factory or contractor facility. Obtain and study system documentation. Discuss design criteria with contractors, SMEs, users, maintainers, etc. The smarter you are about the system, the better decisions you will make, and others will not be able to hide issues that need to be brought out into the open.

- Early operational test agency (OTA) involvement is critical to reducing OT risk. The OTA should/could assist in requirements and concepts of operations (CONOPS) development, in early identification of T&E related systems concerns, in providing input aimed at conducting selected developmental test (DT) events in a more operational environment, etc.

- Plan for proper training, and make the case for extra training time, if needed. Training usually results in a better and safer test, better information, and better decisions. It almost always pays off in the long run.

TEAMWORK, COMMUNICATION (46.8 PERCENT)

Many students cited the importance of teamwork and communication as critical to minimizing and/or preventing T&E problems. Frequent, open, and timely communication (integrated product team [IPT] meetings or otherwise), along with consulting with SMEs when needed, undoubtedly increases the chance of program success. Students had a number of recommendations concerning the importance of teamwork and communication:

- Get support from and educate all involved commands and customers. Early involvement of all key parties (developmental and operational testers, evaluators, user representatives, program office personnel, range and safety personnel, specialty area experts, contractors, etc.) is critical. Resolve conflicts early, if possible. Understand the decision cycles and decision criteria of all involved organizations.

- Consult experts for any important matter, when lacking the necessary expertise. Ask for help when needed!

- Do not punish subordinates for finding errors or performing necessary rework. Establish a healthy team environment where discussions are open and candid and where people do not hide mistakes and problems.

- The IPTs work well when used correctly. They should consist of qualified and empowered team members from all key organizations and stakeholders, plus any needed SMEs. Get the best experts from each organization. Ensure expectations are well understood. If more organizations have input and more points-of-view are considered, more creative options may be generated, which lead to better decisions with better buy-in. There should be consistent, success-oriented, proactive participation: open discussions with no secrets. Issues and concerns should be raised and resolved early through critical dialogue, not just "group think." Reasoned disagreement should occur, with decisions based on reaching consensus, if possible. Ethical decision making is important, with action items worked quickly. Properly functioning IPTs can reduce confusion in an already complex process.

- The T&E effort on joint programs is much tougher to coordinate. Establish a joint T&E working integrated process team (WIPT), with the best experts from each service and continuous "up the line" communications. Joint programs are more challenged by rice bowls and politics, so communication is even more critical. Leadership and joint processes need to be established early. Goals, schedules, performance levels, logistics issues, and CONOPS are unique and different for each service. All these issues need to be worked out early.

- Test reports need to clearly communicate the facts. Write reports for all audience levels (executive summary in lay terms for executives, common technical terms for managers, attachments with appropriate technical jargon for engineers and analysts). Not all deficiencies are equal; prioritize and sort boulders from gravel, based on mission impact. Whenever a deficiency is mentioned, address impact and ease of correction. Include charts and tables that are easy to understand. Put the test methodology in an appendix. Report all results, good and bad, and document the value-added of this testing. Report test results with respect to conditions and mechanisms. Report bottom line results—what worked and what did not—and what decision makers need to know concerning the system's mission capabilities and limitations. Bad news does not get better with age; consider a quick-look report or interim results if decision makers need to quickly know the results.

FUNDING, BUDGET, AND COST (46.8 PERCENT)

This group of lessons dealt with the importance of adequate funding and the negative effects of inadequate funding on T&E programs. Student comments are as follows:

- With the current DoD budget situation, one of the hardest hit areas is T&E. But money and time saved by cutting corners up front is invariably spent in fixing problems later. Customers are unhappy, and it leads to increased cost in the long-run.

- Because of funding limitations, testers commonly execute only a small subset of the test events that should be required. Lack of funding results in delayed testing, test events that are limited in scope, data or reports that are not delivered, cancelled test events, and/or a nonrobust test program that fails to find the critical problems or issues. Unforeseen test requirements sometimes arise for which funding is unavailable.

- Seek to document and educate management and leadership that cost savings obtained by reduced testing may compromise or jeopardize final system quality or operation. And reduced testing, which allows systems to be fielded with undetected problems, could potentially endanger lives.

- To mitigate the risk of inadequate funding, seek to determine realistic cost estimates early in program development. Conduct thorough budget planning and review, consult with experts as needed, and include adequate management reserve for unanticipated problems. If there are cost and schedule constraints, fully document the impact. Include risk analysis and cost/benefit analysis to prioritize limited resources.

Testing is often hindered because of inadequate time available for testing.

SCHEDULE (40.5 PERCENT)

This group of findings dealt with the importance of an adequate schedule and the negative effects of inadequate schedule on T&E programs. Testing is often hindered because of inadequate time available for testing. Overly optimistic schedule estimates commonly lead to this problem as well as unforeseen problems, which decrease the time available or increase the time needed for testing. Program managers sometimes curtail testing in order to make up for lost schedule time. Students said the following:

- Take the time to understand and develop a realistic plan and schedule, including consulting with experts as necessary. Plan for realistic test and program schedules,

with time allowed for things such as maintenance, bad weather, and crew rest. Add cost and schedule contingency to each activity, not just at the end.

- Communicate the need for adequate scheduling, including probable effects if needed or planned testing is delayed or cancelled. A limited test program often results in higher overall program costs and longer overall schedule because discovery of problems is delayed. If there are schedule constraints, fully document the impact. Include risk analysis and cost/benefit analysis to prioritize limited resources.

COMPARISON WITH OTHER STUDIES

A literature search was conducted, and the results from this study were compared with the results from five other similar studies and documents. These studies and documents all examined and/or presented T&E best practices and lessons learned. A short description of each study and document follows:

DEFENSE ACQUISITION GUIDEBOOK SECTION 9.8, BEST PRACTICES (DAG, 2004)

This list of T&E best practices was prepared by developmental test and evaluation (DT&E), operational test and evaluation (OT&E), and live fire test and evaluation (LFT&E) experts at the Office of the Secretary of Defense (OSD). Best practices are offered to increase the likelihood of a successful T&E program. Some commercial industry T&E best practices are included in the list. However, other than that, the methodology for developing the DAG list of best practices is not stated. Since the OSD experts (who developed DAG Section 9.8) have extensive knowledge and experience in T&E oversight of DoD acquisition programs, presumably the list of best practices is (at least partially) based on this extensive T&E knowledge, experience, and expertise.

A MORE CONSTRUCTIVE TEST APPROACH IS KEY TO BETTER WEAPON SYSTEM OUTCOMES (GAO, 2000)

In this report, the General Accounting Office (GAO) examines (a) how the conduct of T&E affects commercial and DoD program outcomes, (b) how best commercial T&E practices compare with DoD's, and (c) what factors account for the differences in these practices. The report includes detailed discussion of DoD and commercial T&E best practices and lessons learned, along with recommendations for improving the conduct of T&E within DoD. The GAO conducted literature searches, interviewed numerous T&E experts, examined four DoD weapon programs, and analyzed T&E best practices of five leading commercial firms, including site visits with structured interview questions sent in advance of each visit.

RECURRING LESSONS IN WEAPON T&E PROGRAMS (HOIVIK, 2000)

This article summarizes and discusses significant issues and problem areas in conducting DoD T&E programs. More detailed analyses and findings may be found in two Naval Postgraduate School Master of Science theses, which are referenced in the article. Sources for the research efforts included studies of T&E in various major system acquisition programs, including information from program office personnel, testers, analysts, user representatives, and contractor T&E personnel. Literature searches and reviews were also conducted.

A STUDY OF COMMERCIAL INDUSTRY BEST PRACTICES IN TEST & EVALUATION WHICH ARE POTENTIALLY APPLICABLE TO DOD DEVELOPMENTAL TEST AND EVALUATION (SAIC, 2002)

This study presents a detailed discussion of commercial industry best practices, including how and why these T&E best practices have led to industry success. The study team made site visits to 12 leading commercial firms and asked them to identify T&E best practices that make them successful (structured interview questions were sent in advance of each visit). The team met with senior corporate managers, engineers, and technicians. The four focus areas for gathering information were: philosophy, policy, and approach; test investment; test execution; and test evaluation.

FLOYD AND WALLY'S OPERATIONAL TEST AND EVALUATION TOP 10 LESSONS LEARNED (SMITH & TUBELL, 2001)

Using knowledge acquired from their direct and indirect experience in Army OT&E, the authors share and discuss their hard-won lessons learned. Their top 10 lessons learned are presented, along with advice and recommendations.

COMPARISON WITH OTHER STUDIES: FINDINGS

Table 3 compares the results discussed in this article with results from the above five studies and documents. To develop Table 3, it was determined which of the 18 lessons learned/best practice categories were listed and/or discussed in each of the five documents. The 18 categories, ranked by frequency, represent only the results of the study discussed in this article. The author did not attempt to rank the findings from the other five documents. The author also did not address information contained in the other five documents that was outside the scope of the 18 categories.

The research methodology and/or population studied were different in each of the above studies. But the findings were similar in that this article's top findings were discussed in all of the studies, whereas the lower categories showed up in less of the studies (or in this article's study only). It is significant that the top six areas were discussed in all of the studies, as this shows that the same sorts of issues have repeatedly surfaced over the past five to ten years.

TABLE 3.
TEST AND EVALUATION BEST PRACTICES/LESSONS LEARNED STUDIES AND DOCUMENTS

LESSONS LEARNED/ BEST PRACTICE CATEGORY	MY FINDINGS	DAG	GAO	HOIVIK	SAIC	SMITH AND TUBELL
1. Test design, test methods, test execution, and analysis methods	X	X	X	X	X	X
2. Test planning	X	X	X	X	X	X
3. Teamwork and communication	X	X	X	X	X	X
4. Funding, budget, and cost	X	X	X	X	X	X
5. Schedule	X	X	X	X	X	X
6. Test infrastructure, test tools, test articles, and instrumentation	X	X	X	X	X	X
7. Test requirements	X			X	X	X
8. Safety and risk management	X		X	X	X	
9. Government leadership/ management issues, and organizational politics	X	X	X	X		
10. Contractor issues, including contractor leadership and management issues	X				X	
11. Training issues	X	X			X	
12. Modeling and simulation (used in conjunction with testing)	X	X			X	
13. Interfaces, interoperability, and integration issues	X	X			X	
14. “Stuff happens” (Murphy’s law, weather)	X					
15. Manpower issues	X			X	X	
16. Immature technology and/or immature system	X		X			
17. COTS and Non-Development Item (NDI) issues	X			X		
18. Poor judgment	X					

CONCLUSIONS

Despite good intentions and some DoD progress, weapon system programs still suffer from persistent problems associated with late or incomplete testing (GAO, 2000). Several common lessons learned have surfaced over the years, as evidenced by the similar findings from this study and the five other studies and documents.

The purpose of this article is to identify and discuss common T&E best practices and lessons learned, thereby enabling possible cost and schedule savings and improved test results from future T&E efforts. Testers, evaluators, and program office personnel can certainly benefit from applying these lessons where they are not already doing so. Decision makers and acquisition/program leadership can benefit by better understanding the top T&E related issues, as reported by field-level T&E personnel.

Karen M. Stadler has 17 years of engineering/acquisition experience in the T&E; systems planning, research development, and engineering (SPRDE); and production, quality, and manufacturing (PQM) career fields. She is currently a professor of T&E with the Defense Acquisition University (DAU), and teaches all DAU T&E classes and several DAU systems engineering classes. She has a B.S. in chemical engineering from Purdue University, and is a former U.S. Naval officer.

(E-mail address: karen.stadler@dau.mil)

AUTHOR **BIOGRAPHY**

REFERENCES

Department of Defense (DoD). (2004, November). *Defense acquisition guidebook* (DAG 9.8). Washington, DC: Author.

General Accounting Office (GAO). (2000, July). *A more constructive test approach is key to better weapon system outcomes* (GAO/NSIAD-00-199). Washington, DC: Author.

Hoivik, T. H. (2000, September/October). Recurring lessons in weapon T&E programs. *Army AL&T*, 35–36.

Science Applications International Corporation (SAIC). (2002, July). *A study of commercial industry best practices in test & evaluation which are potentially applicable to DoD developmental test and evaluation.* Washington, DC: Author.

Smith, F. B., & Tubell, W. (2001, May/June). Floyd and Wally's operational test and evaluation top 10 lessons learned. *Army AL&T*, 26–27.