A Study of Best Practice Adoption by Defense Acquisition Programs

Dr. Richard Turner
The George Washington University

Tuesday, 30 April 2002
Track 1: 3:00 - 3:40
Ballroom A

Best practices are often recommended as a way to improve the success and quality of software-intensive system acquisitions within the Department of Defense. However, one of the primary questions raised by this recommendation is, "To what degree have existing projects, in fact, adopted the best practices?" While the assumption is that most programs are not using best practices, there are few specific studies available. This article presents the results of a study that measured best practice adoption in defense acquisitions.

The United States Department of Defense (DoD) spends an estimated $20 billion a year on software to support its infrastructure; operate its weapons systems; and provide command, control, communications, computing, intelligence, surveillance, and reconnaissance functions. DoD acquires the large majority of this software from contractor sources.

There have been significant cost overruns and schedule delays experienced in DoD software-intensive system acquisitions, resulting in numerous audits and evaluations of acquisition programs by independent government and industry organizations. Such evaluations have consistently indicated that programs are at risk partially because of failure to implement best practices. The evaluations have recommended the implementation of a variety of practices to improve performance [1-8].

A study by Anderson and Rebentisch [9] of 23 military programs found that practice implementation for eight recommended commercial practices[1] ranged from 17 percent to 83 percent with essentially half of the practices implemented under 40 percent of the time and half over 50 percent of the time.

This article reports on research conducted to estimate how broadly acquisition best practices are implemented within DoD. The study involved developing and conducting a survey to establish the implementation and perceived effectiveness of a set of best practices.

Study Methods

The first critical issue was deciding whom to survey. It was desirable to obtain as wide a sample as possible with the least amount of interference in the acquisition program activities. For this reason, it was determined that the most effective way to access a wide variety of projects was to contact the various military software centers that provide software expertise to the programs. These centers act as intramilitary consultants or centers of excellence to provide expert resources to the acquisition program offices. Their personnel are in a position to provide informed judgments without any political bias from program loyalty.

> *"... evaluations have consistently indicated that programs are at risk partially because of failure to implement best practices."*

The second critical decision was selecting the best practices for evaluating adoption. Since one of the objectives of this research was to determine how to support the implementation of best practices, it was decided to use the most widely known and oldest set of practices as the baseline for the adoption study. This would maximize the odds that program managers would have heard of the practices and that the acquisition personnel would have encountered them in use. Therefore, the original nine Airlie practices[2] were used in the survey.

The following definitions of the Airlie practices are derived from the Software Program Managers Network (SPMN) materials [10, 11]. Copies of the SPMN material defining the practices were included in the materials sent to the survey participants. They are as follows:

1. Formal Risk Management. A formal risk management process requires acceptance of risk as a major consideration; commitment of program resources to managing risk; and use of planned, documented methods for identifying, monitoring, and managing risks.
2. Agreement on Interfaces. A baseline interface specification must be established and agreed to by all stakeholders before implementation activities begin. A separate software specification must be developed with explicit and complete interface information. This is particularly critical with human/machine interfaces and where system interoperability is a requirement.
3. Formal Inspections. Inspections of all acquisition and development documentation should be conducted according to planned, documented processes and the results placed under configuration control, tracked, and resolved.
4. Metrics-Based Scheduling and Management. Statistical quality control of costs and schedules should be maintained. Reasonable cost and schedule projections should be made before program start, and specific measurement processes should be put in place early in the program and rigorously followed. Measurement results should figure prominently in program reviews and management decisions.
5. Binary Quality Gates at the Inch-Pebble Level. Status should be tracked through binary completion of relatively small tasks. Activities are either incomplete or complete. This is to prevent the "80-percent-complete" syndrome where the estimated completion figure is reported without particular rigor.
6. Program-Wide Visibility of Progress vs. Plan. Core indicators of project health and performance should be readily available to all project participants. Anonymous feedback channels

should be provided to enable bad news to be propagated up and down the project hierarchy without fear of reprisal for truth telling.

7. Defect Tracking Against Quality Targets. Defects should be tracked according to a planned, documented process; measured against established targets; and systematically tracked through removal or resolution.
8. Configuration Management. A planned, documented process is followed to identify, document, monitor, evaluate, control, and approve changes made during the system life cycle to any system-related artifact that is shared by more than one individual or organization.
9. People-Aware Management Accountability. Management should treat personnel as their principle resource by staffing qualified people, encouraging continuous improvement, and fostering an environment conducive to low voluntary personnel turnover.

Survey Instrument

In developing the instrument, we found it useful to think of adoption as having two components: awareness and implementation. Awareness, as defined by Hilburn [12], represents a level of individual knowledge about the practice that includes the following:

- Understanding of the existence and context of the practice within the context of software acquisition.
- A general, informal explanation of the practice.
- Identification of references (human/written) that provides greater depth of knowledge about the practice.

The other component, implementation, requires that the organization put into place the requisite infrastructure, training, resources, and policy to effectively utilize the practice in doing business.

The adoption survey instrument was designed to provide data that addressed both awareness and adoption. It collected data in the following areas:

1. The number of programs supported by the respondent to establish the overall program sample.
2. The size of those programs as designated by Acquisition Category (ACAT): ACAT I, ACAT II, ACAT III, or Other.[3]
3. The quality of practice adoption for each program as measured by the compliance with the practice definition in the Airlie material. This was captured by having each participant differentiate between full compliance and partial compliance for each of the projects. Partial compliance would be a surrogate for *awareness*, while full compliance would represent *implementation*. Obviously, this is not a perfect surrogate. However, since partial compliance does, in fact, capture awareness, the worst error would be that of underestimating awareness. It was decided that such an error could be dealt with by considering it in the analysis. The data were gathered for each Airlie practice as the following:
   - The number of programs in each size category that fully implemented the practice.
   - The number of programs in each size category that implemented some facet of the practice.
4. An evaluation of the perceived overall effectiveness of the practice as observed by the center personnel measured on a five-point scale: Highly Effective (5), Very Effective (4), Moderately Effective (3), Somewhat Effective (2), Negligibly Effective (1). This scale was chosen to reflect that as best practices, the practices were, at worse, ineffective.

Service	ACAT I	ACAT II	ACAT III	Other	Total
Air Force	1	0	16	62	**79**
Army	9	11	13	6	**39**
Navy	2	6	7	17	**32**
Total	**12**	**17**	**36**	**85**	**150**

Table 1: *Programs Reported by Service and ACAT Designation*

ACAT I	ACAT II	ACAT III	Other
8%	11%	24%	57%

Table 2: *Percentage of Programs Represented by ACAT Designation*

Air Force	Army	Navy
53%	26%	21%

Table 3: *Percentage of Programs Represented by Service*

Results

Of the 14 centers asked to participate, seven responded to the data call. Six of the seven provided the requested data:

1. Army TACOM TARDEC
2. Army CECOM
3. Navy NAVSEA
4. Navy NUWC
5. Navy NAVAIR
6. Air Force ESC

The seventh respondent, Air Force ASC, provided narrative comments only. The responses covered 150 software acquisition programs broken out as shown in Tables 1 through 3.

Airlie Practice Adoption Data

The responses to the survey resulted in 1,350 possible program-practice pairs (150 programs times nine practices) where a particular Airlie practice could be adopted by a particular program. Table 4 shows the summarized results of the survey when calculated against this full complement of program-practice pairs. The terms *partial* and *full* refer to whether the respondent indicated that the program partially or fully implemented the practice.

Table 5 presents the adoption data by practice. The percentages represent the

Table 4: *Overall Results of Adoption Study (Percent of Possible Program/Practice Pairs)*

Overall	ACAT I	ACAT II	ACAT III	Other	Total
Partial	67%	52%	38%	69%	**59%**
Full	13%	29%	31%	24%	**25%**
Total	**80%**	**80%**	**69%**	**93%**	**84%**
Air Force	**ACAT I**	**ACAT II**	**ACAT III**	**Other**	**Total**
Partial	44%	None	56%	78%	**73%**
Full	56%	None	12%	21%	**20%**
Total	**100%**	**None**	**68%**	**99%**	**93%**
Army	**ACAT I**	**ACAT II**	**ACAT III**	**Other**	**Total**
Partial	64%	45%	21%	13%	**36%**
Full	9%	24%	50%	67%	**36%**
Total	**73%**	**70%**	**70%**	**80%**	**72%**
Navy	**ACAT I**	**ACAT II**	**ACAT III**	**Other**	**Total**
Partial	89%	63%	29%	54%	**52%**
Full	11%	37%	40%	18%	**26%**
Total	**100%**	**100%**	**68%**	**72%**	**78%**
Note: Numbers may not sum due to rounding					

Best Practice	ACAT I		ACAT II		ACAT III		Other	
	Partial	Full	Partial	Full	Partial	Full	Partial	Full
Formal risk management	92%	8%	41%	29%	39%	31%	72%	19%
Agreement on interfaces	75%	25%	41%	59%	47%	36%	67%	29%
Formal inspections	50%	8%	71%	0%	36%	28%	65%	20%
Metric-based scheduling and management	83%	8%	59%	12%	44%	22%	69%	19%
Binary quality gates at the inch-pebble level	33%	0%	65%	6%	14%	22%	71%	16%
Program-wide visibility of progress vs. plan	83%	8%	47%	53%	47%	36%	72%	26%
Defect tracking against quality targets	25%	17%	29%	41%	31%	36%	68%	31%
Configuration management	67%	33%	47%	53%	42%	42%	66%	34%
People-aware management accountability	92%	8%	65%	6%	42%	25%	68%	21%

Table 5: *Adoption Results by Practice*

Best Practice	Effectiveness*					Effectiveness Score*
	HE	VE	ME	SE	NE	
Formal risk management	14%	**50%**	29%	7%	0%	3.71
Agreement on interfaces	36%	**43%**	21%	0%	0%	4.14
Formal inspections	21%	**36%**	14%	21%	7%	3.43
Metric-based scheduling and management	15%	38%	**46%**	0%	0%	3.69
Binary quality gates at the inch-pebble level	18%	18%	**36%**	27%	0%	3.27
Program-wide visibility of progress vs. plan	15%	**69%**	8%	8%	0%	3.92
Defect tracking against quality targets	23%	**38%**	**38%**	0%	0%	3.85
Configuration management	33%	**47%**	20%	0%	0%	4.13
People-aware management accountability	14%	21%	**43%**	21%	0%	3.29

*5=Highly Effective (HE), 4=Very Effective (VE), 3=Moderately Effective (ME), 2=Somewhat Effective (SE), 1=Negligibly Effective (NE).

Table 6: *Effectiveness Results by Practice*

number of projects reporting the particular value divided by the total number in the ACAT designation category.

Airlie Practice Effectiveness Data

Table 6 presents the effectiveness ratings by practice as percentages of respondents rating the practice in the particular categories from Highly Effective (HE), Very Effective (VE), Moderately Effective (ME), Somewhat Effective (SE), through Negligibly Effective (NE). The effectiveness value is the mean of the scores received using five as the value for HE, four for VE, three for ME, two for SE, and one for NE. Table 7 shows the overall results with the practices ranked by effectiveness.

Common barriers reported were as follows:

- Lack of Resources. In general, program offices lack sufficient software-educated staff to implement many of the practices. Some lack sufficient core staff and are too busy fighting everyday fires to even consider best practices.
- Inadequate Contracts. Some contracts do not allow the flexibility needed to implement practices that require developer support without costly time- and energy-consuming modifications.
- Data Accuracy, Availability, and Latency. Practices often depend on accurate, timely data that are simply not available to many DoD programs. The lack or unavailability of valid, useful historical data often stymies estimation practices.
- Management Awareness and Commitment. Several comments were made about managers who did not understand that there was a problem or who did not want to spend the resources on a practice that might actually add risk rather than reduce it. Management must be willing to expend resources and perhaps political capital to institute practices.
- Lack of Credible Evidence. Comments were received concerning the need to prove that the benefits of practices were worth the cost of practice implementation.

Table 7: *Overall Adoption and Effectiveness (Practices in Order of Effectiveness Ranking)*

Overall	P+F*	F* Only	Effectiveness**
Agreement on interfaces	94%	34%	4.14
Configuration management	96%	38%	4.13
Programwide visibility of progress vs. plan	94%	30%	3.92
Defect tracking against quality targets	83%	32%	3.85
Formal risk management	84%	22%	3.71
Metric-based scheduling and management	81%	18%	3.69
Formal inspections	76%	19%	3.43
People-aware management accountability	83%	19%	3.29
Binary quality gates at the inch-pebble level	69%	15%	3.27

*P+F = Partial and full implementations combined; F Only = Full implementations only.
**5=Highly Effective, 4=Very Effective, 3=Moderately Effective, 2=Somewhat Effective, 1=Negligibly Effective.

Analysis

The first observation is that the practices are widely recognized across the programs as indicated by the high percentage (84 percent) of either partial or full implementations. Any particular practice was implemented in some form for at least 69 percent of the projects reporting. Three of the practices were implemented in around 95 percent of the projects. However, when only considering full implementations, those figures drop dramatically to an average of 25 percent across all programs with the lowest rate for a specific practice adoption of 15 percent and the highest rate of 38 percent.

Across services, adoption rates were generally consistent. The Army had a generally higher percentage of full implementation. Figure 1 illustrates the relationships between the services.

The practices were evaluated as relatively effective, with the majority of responses falling in the *moderately effective* to *very effective* range. The practice considered least effective (binary quality gates) still had a mean score of 3.27, placing it in the moderately effective category.

If we look at the adoption and effectiveness data together, there is some correlation between the perceived effectiveness of a practice and its adoption. This seems logical, since in budget- and schedule-constrained programs, the practices with the highest effectiveness would seem more likely to be implemented.

Conclusions

As stated earlier, we have used partial compliance as a surrogate that implies awareness and full compliance to imply implementation. Therefore, the primary finding from the adoption research is that despite the widespread awareness of the best practices (the average of programs implementing practices was 85 percent, which as we

previously noted, is probably an understatement of true awareness), there is very little actual implementation (average 25 percent). If we assume that the practice must be fully implemented to gain substantial benefit, little value is being realized.

In general, full implementation is not required; however, when coupled with the environment of defense acquisition, full (and possibly formal) implementation is the only way that a practice can expect to maintain focus and resources long enough to achieve benefits. This is particularly true of practices that have longer benefit latency.

Judging by the effectiveness ratings, the Airlie practices have stood the test of time and represent valid best practices. Within the Airlie practices, configuration management, agreement on interfaces, and risk management are essentially fundamental project management activities.

As one respondent pointed out, "A number of the things promoted by the SPMN are simply established good practices that were known and practiced before the Airlie group documented them as best practices." That said, there are still many programs that do not implement these practices effectively and so should be reminded of their importance.

The research as conducted describes an environment where managers are aware of the benefits of acquisition practices, but they do not implement them. Either the barriers that prevent full implementation are sufficiently high to deter action, or the program managers simply choose not to implement the practices. The research supported both of these possibilities.

To reap the benefits of the Airlie practices, or any best practices or other acquisition technology, the software-intensive system acquisition environment needs to be changed. The Software Intensive Systems (SIS) office within the Acquisition Resources and Analysis Directorate of the Under Secretary of Defense for Acquisition, Technology, and Logistics is working to improve policy, transition new acquisition technology into programs, coordinate independent expert program reviews, gather empirical data on best practices, and support broader software-related education across the acquisition workforce.

SIS has completed additional research in the area of best practices in the last year. I will be writing another article that will describe the consolidation of more than 100 published practices into 32 candidate practices. Those practices were evaluated for effectiveness by a panel of experts, and an analysis of their impact on software-intensive, system-acquisition risk areas was performed. Further research on better ways of describing practices in a way more suitable to selection and evaluation by acquisition personnel is in the final stages. The results will be briefed in the SIS track during the 2002 Software Technology Conference. The best practice research will be combined with other SIS efforts, including the TriService Assessment Initiative and the CeBASE Experience Factory pilots, to support better decision making and improved processes in software-intensive system acquisitions across DoD. Further work on documenting and disseminating best practices is being performed in collaboration with the Data and Analysis Center for Software in Rome, N.Y.◆

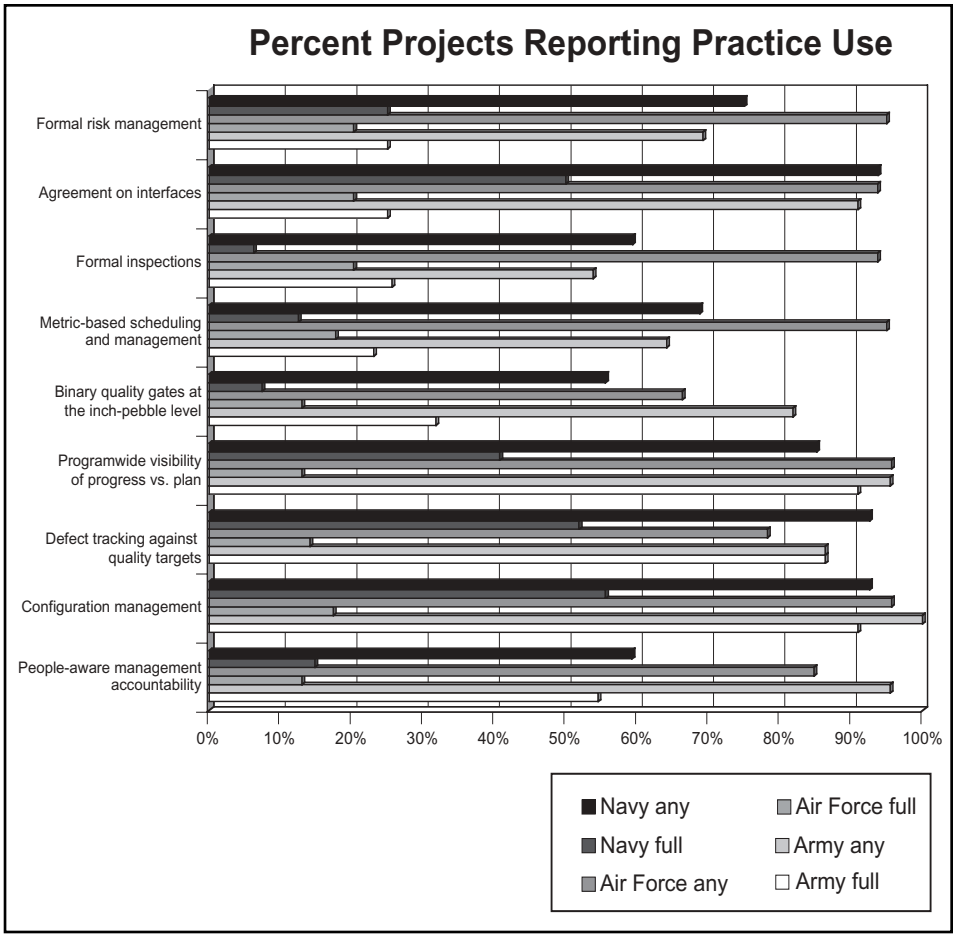

Figure 1: *Comparison of Service Adoption Rates*

References

1. Department of Defense. Defense Software: Review of Defense Report on Software Development Best Practices. Washington, D.C.: General Accounting Office, 2000.
2. Department of Defense. Major Management Challenges and Program Risks: Department of Defense. Washington, D.C.: General Accounting Office, 2001.
3. Department of Defense. Best Practices: Successful Application to Weapon Acquisitions Requires Changes in DoD's Environment. Washington, D.C.: General Accounting Office, 1998.
4. Department of Defense, Office of the Inspector General. Summary of Audits of Acquisition of Information Technology. Washington, D.C.: General Accounting Office, 2000.
5. Charette, R., and J. J. McGarry. Tri-Service Assessment Initiative Phase 1 Systemic Analysis Executive Report. Washington, DC.: Tri-Service Assessment Initiative, 2001.
6. Department of Defense, Defense Science Board. Report of the Defense Science Board Task Force on Defense Software. Washington, D.C.: 2000.
7. Glennan Jr., T.K., et al. Barriers to Managing Risk in Large Scale Weapons System Development Programs. Santa Monica, Calif.: RAND, 1993: 30.
8. Eslinger, S. Software Acquisition and Software Engineering Best Practices. Los Angeles: The Aerospace Corporation, 1999: 40.
9. Anderson, M.H., and E. Rebentisch. "Commercial Practices – Dilemma or Opportunity?" Program Manager 27.2 (1998): 16-21.
10. Software Program Managers Network. "Program Managers Guide to Software Acquisition Best Practices." Ver. 5.

Integrated Computer Engineering 1999.
11. Software Program Managers Network. "16 Critical Software Practices For Performance-based Management." Ver. 5.1. Integrated Computer Engineering 1999.
12. Hilburn, T., et al. FAA-ARA Software Engineering Competency Study: Final Report. Daytona Beach, Fla.: Embry-Riddle Aeronautical University, 1998.

Notes

1. The eight practices studied were past performance, best value, commercial warranties, government/contractor relationship, performance-based specifications, commercial specifications and standards, streamlined contract administration, and use of commercial off-the-shelf/non-developmental item components. The study also found considerable benefits with few drawbacks for using the practices.
2. The nine Airlie Practices were established in 1995 by a group of experts convened by the Navy's Software Program Manager's Network at the Airlie Center outside Warrenton, Va. (now historically referred to as "The Airlie Council").
3. ACAT I is defined as an acquisition program that is not a highly sensitive classified program, and that is designated as a Major Defense Acquisition Program; or estimated to require an eventual total expenditure for research, development, test, and evaluation of more than $355 million in fiscal year (FY) 1996 constant dollars; or for procurement of more than $2.135 billion in FY 1996 constant dollars. ACAT II is defined as those acquisition programs that do not meet the criteria for an ACAT I program, but are estimated to require an eventual total expenditure for research, development, test, and evaluation of more than $135 million in FY 1996 constant dollars; or for procurement of more than $640 million in FY 1996 constant dollars; or if designated as major by the DoD component head. ACAT III is defined as those acquisition programs that do not meet the criteria for an ACAT I or ACAT II. Other is defined as any acquisitions not designated with an ACAT level.

About the Author

Richard Turner, D. Sc., is a member of the Engineering Management and Systems Engineering Faculty at The George Washington University in Washington, DC. Currently, Dr. Turner is the assistant deputy director for Software Engineering and Acquisition in the Software Intensive Systems Office of the Under Secretary of Defense (Acquisition, Technology & Logistics). Dr. Turner is co-author of the book *CMMI Distilled* published by Addison Wesley.

1931 Jefferson Davis Highway
Ste. 104
Arlington, VA 22202
Phone: (703) 602-0851 ext. 124
E-mail: rich.turner@osd.mil