Early Identification of SE-Related Program Risks

Barry Boehm, Dan Ingold, University of Southern California

Kathleen Dangle, Fraunhofer-Maryland

Rich Turner, Stevens Institute of Technology

Paul Componation, University of Alabama-Huntsville

Abstract

This paper summarizes the results of a DoD Systems Engineering Research Center (SERC) project to synthesize analyses of DoD SE effectiveness risk sources into a lean framework and toolset for early identification of SE-related program risks. It includes concepts of operation which enable project sponsors and performers to agree on the nature and use of more effective evidence-based reviews. These enable early detection of missing SE capabilities or personnel competencies with respect to a framework of Goals, Critical Success Factors (CSFs), and Questions determined from leading DoD early-SE CSF analyses. The SE Effectiveness Measurement (EM) tools enable risk-based prioritization of corrective actions, as shortfalls in evidence for each question are early uncertainties, which when combined with the relative system impact of a negative answer to the question, translates into the degree of risk that needs to be managed to avoid system overruns and incomplete deliveries.

Introduction; Motivation and Context

DoD programs need effective systems engineering (SE) to succeed.

DoD program managers need early warning of any risks to achieving effective SE.

This SERC project has synthesized analyses of DoD SE effectiveness risk sources into a lean framework and toolset for early identification of SE-related program risks.

Three important points need to be made about these risks.

- They are generally not indicators of "bad SE." Although SE can be done badly, more often the risks are consequences of inadequate program funding (SE is the first victim of an underbudgeted program), of misguided contract provisions (when a program manager is faced with the choice between allocating limited SE resources toward producing contract-incentivized functional specifications vs. addressing key performance parameter risks, the path of least resistance is to obey the contract), or of management temptations to show early progress on the easy parts while deferring the hard parts till later.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
19 NOV 2009		**00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Early Identification of SE-Related Program Risks	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Systems Engineering Research Institute,Stevens Institute of Technology,1 Castle Point on Hudson,Hoboken,NJ,07030	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

Proceedings of the Conference on Systems Engineering Research, Hoboken, NJ, March 2010. DOD-sponsored. U.S. Government or Federal Rights License

14. ABSTRACT

This paper summarizes the results of a DoD Systems Engineering Research Center (SERC) project to synthesize analyses of DoD SE effectiveness risk sources into a lean framework and toolset for early identification of SE-related program risks. It includes concepts of operation which enable project sponsors and performers to agree on the nature and use of more effective evidence-based reviews. These enable early detection of missing SE capabilities or personnel competencies with respect to a framework of Goals, Critical Success Factors (CSFs), and Questions determined from leading DoD early-SE CSF analyses. The SE Effectiveness Measurement (EM) tools enable risk-based prioritization of corrective actions, as shortfalls in evidence for each question are early uncertainties, which when combined with the relative system impact of a negative answer to the question, translates into the degree of risk that needs to be managed to avoid system overruns and incomplete deliveries.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**15**	

- Analyses have shown that unaddressed risk generally leads to serious budget and schedule overruns.

- Risks are not necessarily bad. If an early capability is needed, and the risky solution has been shown to be superior to the alternatives, accepting and focusing on mitigating the risk is generally better than waiting for a better alternative to show up.

Unlike traditional schedule-based and event-based reviews, the SERC SE EM technology enables sponsors and performers to agree on the nature and use of more effective evidence-based reviews. These enable early detection of missing SE capabilities or personnel competencies with respect to a framework of Goals, Critical Success Factors (CSFs), and Questions determined by the EM task from the leading DoD early-SE CSF analyses. The EM tools enable risk-based prioritization of corrective actions, as shortfalls in evidence for each question are early uncertainties, which when combined with the relative system impact of a negative answer to the question, translates into the degree of risk that needs to be managed to avoid system overruns and incomplete deliveries.

The EM tools' definition of "SE effectiveness" is taken from the INCOSE definition of SE as "an interdisciplinary approach and means to enable the realization of successful systems." Based on this definition, the SERC project proceeded to identify and organize a framework of SE effectiveness measures (EMs) that could be used to assess the evidence that a MDAP's SE approach, current results, and personnel competencies were sufficiently strong to enable program success. Another component of the research was to formulate operational concepts that would enable MDAP sponsors and performers to use the EMs as the basis of collaborative formulation, scoping, planning, and monitoring of the program's SE activities, and to use the monitoring results to steer the program toward the achievement of feasible SE solutions.

Technical Approach

The EM research project reviewed over two-dozen sources of candidate SE EMs, and converged on the strongest sources to be used to identify candidate SE EMs. We developed a coverage matrix to determine the envelope of candidate EMs, and the strength of consensus on each candidate EM. It fed the results back to the source originators to validate the coverage matrix results. This resulted in further insights and added candidate EMs to be incorporated into an SE Performance Risk Framework. The resulting framework is organized into a hierarchy with 4 Goals, 18 Critical Success Factors, and 74 Questions that appeared to cover the central core of common SE performance determinants of SE effectiveness.

Concurrently, the research project was extended to also assess SE personnel competency as a determinant of program success. We analyzed an additional six personnel competency risk frameworks and sets of questions. Their Goals and Critical Success Factors were very similar to those used in the SE Performance Risk Framework, although the Questions were different. The resulting SE Competency Risk Framework added one further Goal of Professional and Interpersonal Skills with five Critical Success Factors, resulting in a framework of 5 Goals, 23 Critical Success Factors, and 81 Questions.

Our initial research focused on identifying methods that might be suitable for assessing the effectiveness of systems engineering on major defense acquisition programs (MDAPs). A literature review identified eight candidate measurement methods: the NRC Pre-Milestone A & Early-Phase SysE top-20 checklist [20]; the Air Force Probability of Program Success (PoPS) Framework [1]; the INCOSE/LMCO/MIT Leading Indicators [24]; the Stevens Leading Indicators (new; using SADB root causes) [34]; the USC Anchor Point Feasibility Evidence criteria [31]; the UAH teaming theories criteria [14]; the NDIA/SEI capability/challenge criteria [15]; and the SISAIG Early Warning Indicators [9] incorporated into the USC Macro Risk Tool [33].

Pages 5-8 of the NRC report [20] suggests a "Pre-Milestone A/B Checklist" for judging the successful completion of early-phase systems engineering. Using this checklist as a concise starting point, we identified similar key elements in each of the other candidate measurement methods, resulting in a coverage matrix with a list of 45 characteristics of effective systems engineering. Figure 1 shows the first page of the coverage matrix. We then had the originators of the measurement methods indicate where they felt the coverage matrix was inaccurate or incomplete. This assessment also identified another six EM characteristics not previously noted.

Figure 1. EM Coverage Matrix

SERC EM Task Coverage Matrix V1.0

	NRC	Probability of Success	SE Leading Indicators	LIPSF (Stevens)	Anchoring SW Process (USC)	PSSES (U. of Alabama)	SSEE (CMU/SEI)	Macro Risk Model/Tool
Concept Dev								
Atleast 2 alternatives have been evaluated	X			x	x	x (w.r.t NPR)	(x)	
Can an initial capability be achieved within the time that the key program leaders are expected to remain engaged in their current jobs (normally less than 5 years or so after Milestone B)? If this is not possible for a complex major development program, can critical subsystems, or at least a key subset of them, be demonstrated within that time frame?	X		(x)	x	x (5 years is not explicitly stated)		(x) (seems to be inferrable from the conclusions)	(x) (implies this)
Will risky new technology mature before B? Is there a risk mitigation plan?	x	x	x		(x)		x	x
Have external interface complexities been identified and minimized? Is there a plan to mitigate their risks?	x		x		x	x	x	x
KPP and CONOPS								
At Milestone A, have the KPPs been identified in clear, comprehensive, concise terms that are understandable to the users of the system?	x	(x)	x	(x)	x (strongly implied)	(x) (implied)	x	x
At Milestone B, are the major system-level requirements (including all KPPs) defined sufficiently to provide a stable basis for the development through IOC?	x	x	(x)	x	x	(x)	(x) (There is no direct reference to this but is inferrable)	x
Has a CONOPS been developed showing that the system can be operated to handle the expected throughput and meet response time requirements?	x	x	(x)	(x)	x	(x) (there is a mention of a physical solution. That's the closest in this regard)	x	x

Legend:
x = covered by EM
(x) = partially covered (unless stated otherwise)

Previous research by the USC team into a macro-risk model for large-scale projects had resulted in a taxonomy of high-level goals and supporting critical success factors (CSFs) based on [28]. This was identified as a potential framework for organizing the 51 EM characteristics identified above. Analysis of the characteristics showed that they could be similarly organized into a series of four high-level goals,

each containing 4-5 CSFs, as seen in Figure 2. Our survey of the existing literature suggests that these CSFs are among the factors that are most critical to successful SE, and that the degree to which the SE function in a program satisfies these CSFs is a measure of SE effectiveness.

Figure 2. Goals and CSFs for SE Performance

High-level Goals	Critical Success Factors
Concurrent definition of system requirements & solutions	Understanding of stakeholder needs
	Concurrent exploration of solutions
	System scoping & requirements definition
	Prioritization/allocation of requirements
System life-cycle organization, planning & staffing	Establishment of stakeholder RAAs
	Establishment of IPT RAAs
	Establishment of resources to meet objectives
	Establishment of selection/contracting/incentives
	Assurance of necessary personnel competencies
Technology maturing & architecting	COTS/NDI evaluation, selection, validation
	Life-cycle architecture definition & validation
	Use of prototypes, models, etc. to validate maturity
	Validated budgets & schedules
Evidence-based progress monitoring & commitment reviews	Monitoring of system definition
	Monitoring of feasibility evidence development
	Monitoring/assessment/re-planning for changes
	Identification and mitigation for feasibility risks
	Reviews to ensure stakeholder commitment

Related to the effectiveness measures of SE performance is the need to measure the effectiveness of the staff assigned to the SE function. Besides the eight SEPRT sources, six additional sources were reviewed for contributions to Personnel Competency evidence questions: the Office of the Director of

National Intelligence (ODNI), Subdirectory Data Collection Tool: Systems Engineering [22]; the INCOSE Systems Engineering Handbook, August 2007 [17]; the ASN (RD&A), Guidebook for Acquisition of Naval Software Intensive Systems, September 2008 [3]; the CMU/SEI, Models for Evaluating and Improving Architecture Competence report [4]; the NASA Office of the Chief Engineer, NASA Systems Engineering Behavior Study, October 2008 [34]; and the National Research Council, Human-System Integration in the System Development Process report, 2007 [23].

These were analyzed for candidate knowledge, skills, and abilities (KSA) attributes proposed for systems engineers. Organizing these work activities and KSAs revealed that the first four goals and their CSFs were in common with the EM taxonomy. This is shown in Figure 3, which shows the compatibility of the four goals in the EM taxonomy with the first four goals in the National Defense Industry Association's SE Personnel Competency framework and those in the CMU/SEI Models for Evaluating and Improving Architecture Competence report.

Figure 3. Comparison of EM Competency Framework with NDIA and SEI Counterparts

SERC EM Framework	NDIA Personnel Competency FW	SEI Architect Competency FW
Concurrent Definition of System Requirements & Solutions	Systems Thinking	Stakeholder Interaction
System Life Cycle Organization, Planning, Staffing	Life Cycle View	Other phases
Technology Maturing and Architecting	SE Technical	Architecting
Evidence-Based Progress Monitoring & Commitment Reviews	SE Technical Management	Management
Professional/ Interpersonal (added)	Professional/ Interpersonal	Leadership, Communication, Interpersonal

As one might expect, the two competency frameworks also had a fifth goal emphasizing professional and interpersonal competencies. Drawing on these and the other Personnel Competency sources cited above, an additional goal and its related CSFs were added for the EM Competency framework, as presented in Figure 4.

Figure 4. Additional goals and CSFs for SE competency

High-level Goal	Critical Success Factors
	Ability to plan, staff, organize, team-build, control, and direct systems engineering teams
	Ability to work with others to negotiate, plan, execute, and coordinate complementary tasks for achieving program objectives
Professional and interpersonal skills	Ability to perform timely, coherent, and concise verbal and written communication
	Ability to deliver on promises and behave ethically
	Ability to cope with uncertainty and unexpected developments, and to seek help and fill relevant knowledge gaps

Question-Level Impact/Evidence Ratings and Project SE Risk Assessment

Using these relatively high-level criteria, however, it is difficult to evaluate whether the SE on a particular program adequately satisfies the CSFs. In its approach to evaluating macro-risk in a program, [31] suggests that a goal-question-metric (GQM) approach [4] provides a method to accomplish this evaluation. Following this example, we developed *questions* to explore each goal and CSF, and devised *metrics* to determine the relevance of each question and the quality of each answer.

The researchers began question development for the SE performance framework with the checklist from [20]. Further questions were adapted from the remaining EM characteristics, rewritten as necessary to express them in the form of a question. Each question is phrased such that, answered affirmatively, it indicates positive support of the corresponding CSF. Thus, the strength of support for each answer is related to the relative *risk probability* associated with the CSF that question explores.

Rather than rely simply on the opinion of the evaluator as to the relative certainty of positive SE performance, a stronger and more quantifiable *evidence-based* approach was selected. The strength of the response is related to the amount of evidence available to support an affirmative answer—the stronger the evidence, the lower the risk probability. Feedback from industry, government, and academic participants in workshops conducted in March and May 2009 suggested that a simple risk probability scale with four discrete values be employed for this purpose.

Evidence takes whatever form is appropriate for the particular question. For example, a simulation model might provide evidence that a particular performance goal can be met. Further, the strongest evidence is that which independent expert evaluators have validated.

Recognizing that each characteristic might be more or less applicable to a particular program being evaluated, the questions are also weighted according to the *risk impact* that failure to address the

question might be expected to have on the program. Again based on workshop feedback, a four-value scale for impact was chosen.

The product of the magnitude of a potential loss (the risk impact) and the likelihood of that loss (the risk probability) is the *risk exposure*. Although risk exposure is generally calculated given quantitative real-number estimates of the magnitude and probabilities of a loss, the assessments of risk impact and risk probability described above use an ordinal scale. However, as shown in the tool below, we have associated quantitative ranges of loss magnitude and loss probability with the rating levels, providing a quantitative basis for a mapping between the four-value risk probability and risk impact scales to a discrete five-value risk exposure scale.

Prototype SE Effectiveness Risk Tools

As a means to test the utility of these characteristics for assessing systems engineering effectiveness, using the GQM approach outlined above, the researchers created prototype tools that might be used to perform periodic evaluations of a project, similar to a tool used in conjunction with the macro-risk model described above. The following section describes this prototype implementation in further detail.

SE Performance Risk Tool

The Systems Engineering Performance Risk Tool (SEPRT) is an Excel spreadsheet-based prototype focused on enabling projects to determine their relative risk exposure due to shortfalls in their SE performance relative to their prioritized project needs. It complements other SE performance effectiveness assessment capabilities such as the INCOSE Leading Indicators, in that it supports periodic assessment of evidence of key SE function performance, as compared to supporting continuous assessment of key project SE quantities such as requirements volatility, change and problem closure times, risk handling, and staffing trends.

The operational concept of the SEPRT tool is to enable project management (generally the Project Manager or his/her designate) to prioritize the relative impact on the particular project of shortfalls in performing the SE task represented in each question. Correspondingly, the tool enables the project systems engineering function (generally the Chief Engineer or Chief Systems Engineer or their designate) to evaluate the evidence that the project has adequately performed that task. This combination of impact and risk assessment enables the tool to estimate the relative project risk exposure for each question, and to display them in a color-coded Red-Yellow-Green form.

These ideas were reviewed in workshops with industry, government, and academic participants conducted in March and May 2009, with respect to usability factors in a real project environment. A consensus emerged that the scale of risk impact and risk probability estimates should be kept simple and easy to understand. Thus a red, yellow, green, and grey scale was suggested to code the risk impact; and a corresponding red, yellow, green, and blue scale to code the risk probability. These scales are discussed in more depth below. An example of the rating scales, questions, and calculated risk exposure in the prototype tool is presented in Figure 5 below.

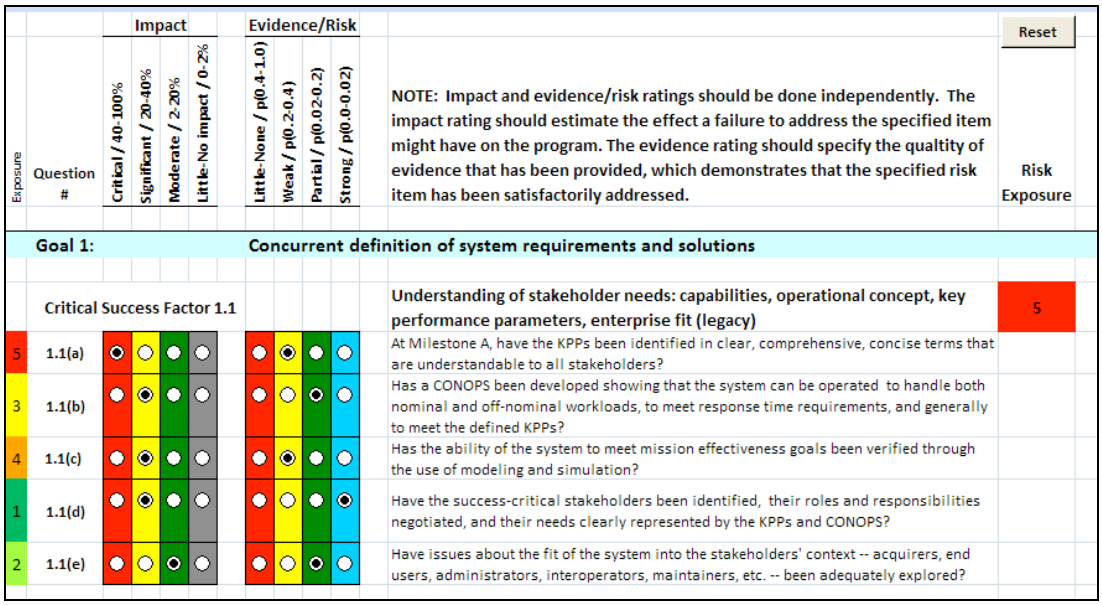


Figure 5. The SEPRT Tool Seeks Performance Evidence

Risk impact ratings vary from a *critical impact* (40-100%; average 70% cost-schedule-capability shortfall) in performing the SE task in question (red) through *significant impact* (20-40%; average 30% shortfall: yellow) and *moderate impact* (2-20%; average 11% shortfall: green) to *little-no impact* (0-2%; average 1% shortfall: gray). These relative impact ratings enable projects to tailor the evaluation to the project's specific situation. Thus, for example, it is easy to "drop" a question by clicking on its "No Impact" button, but also easy to restore it by clicking on a higher impact button. The rating scale for the impact level is based on the user's chosen combination of effects on the project's likely cost overrun, schedule overrun, and missing percent of promised over actual delivered capability (considering there are various tradeoffs among these quantities).

Using Question 1.1(a) from 5 as an example, if the project were a back-room application for base operations with no mission-critical key performance parameters (KPPs), its impact rating would be Little-No impact (Gray). However, if the project were a C4ISR system with several mission-critical KPPs, its rating would be Critical impact (Red).

The Evidence/Risk rating is the project's degree of evidence that each SE effectiveness question is satisfactorily addressed, scored (generally by the project Chief Engineer or Chief Systems Engineer or their designate) on a risk probability scale: the less evidence, the higher the probability of shortfalls. As with the Impact scale, the Evidence scale has associated quantitative ratings: Little or No Evidence: P = 0.4 - 1.0; average 0.7; Weak Evidence: P = 0.2- 0.4; average 0.3; Partial Evidence: P = 0.02 – 0.2; average 0.11; Strong Evidence: P = 0 – 0.02; average 0.01.

Again, using Question 1.1(a) from Figure as an example analyzing a C4ISR system with several mission-critical KPPs, then a lack of evidence (from analysis of current-system shortfalls and/or the use of operational scenarios and prototypes) that its "KPPs had been identified at Milestone A in clear,

comprehensive, concise terms that are understandable to the users of the system" would result in a High risk probability, while strong and externally validated evidence would result in a Very Low risk probability.

Using the average probability and impact values presented above, the average-valued Risk Exposure = P(Risk) * Size(Risk) relative to 100% implied by the ratings is presented in Figure 6. The SEPRT tool provides a customizable mapping of each impact/probability pair to a color-coded risk exposure, based on the above table. For each question, the *risk exposure* level is determined by the combination of risk impact and risk probability, and a corresponding risk exposure color-coding is selected, which ranges from red for the highest risk exposure to green for the lowest. Figure 6 the default color-coding used in the SEPRT tool; an additional Excel sheet in the tool enables users to specify different color codings.

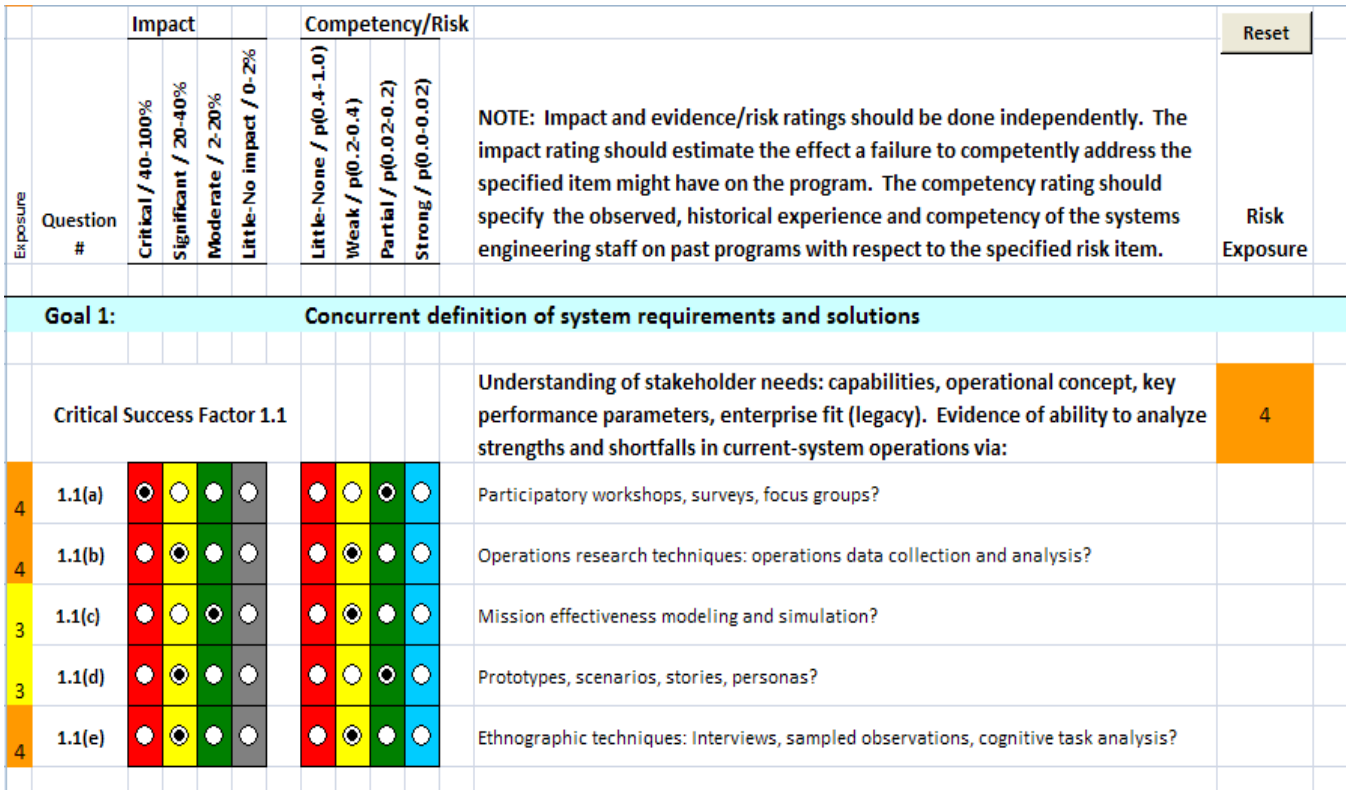
Figure 6. Average risk exposure calculation and default color code

Impact // Probability	Very Low	Low	Medium	High
Critical	0.7	7.7	21	49
Significant	0.3	3.3	7.7	21
Moderate	0.11	1.21	3.3	7.7
Little-No Impact	0.01	0.11	0.3	0.7

As seen in 5, the risk exposure resulting from scoring the impact and risk of each question is presented in the leftmost column. Based on suggestions from workshop participants, the current version of the tool assigns the highest risk exposure level achieved by any of the questions in a CSF as the risk exposure for the overall CSF. This maximum risk exposure presented in the rightmost column for the CSF. This rating method has the advantages of being simple and conservative, but might raise questions if, for example, CSF 1.1 were given a red risk exposure level for one red and four greens, and a yellow risk exposure level for five yellows. Experience from piloting of the tool has suggested refinements to this approach, discussed later in this report.

SE Competency Risk Tool

The initial section of the Systems Engineering Competency Risk Tool (SECRT) is shown in Figure 7. It functions in the same way as the SEPRT tool described above, but its questions address key considerations of personnel competency for each CSF. The space limitations of this paper preclude showing all of the SEPRT and SECRT questions corresponding to the goals and CSF. They are provided in the downloadable tools and SERC EM project Final Technical Report [36] at the SERC web site at **TBD.**



Exposure	Question #	Impact: Critical / 40-100%	Impact: Significant / 20-40%	Impact: Moderate / 2-20%	Impact: Little-No impact / 0-2%	Competency/Risk: Little-None / p(0.4-1.0)	Competency/Risk: Weak / p(0.2-0.4)	Competency/Risk: Partial / p(0.02-0.2)	Competency/Risk: Strong / p(0.0-0.02)	NOTE: Impact and evidence/risk ratings should be done independently. The impact rating should estimate the effect a failure to competently address the specified item might have on the program. The competency rating should specify the observed, historical experience and competency of the systems engineering staff on past programs with respect to the specified risk item.	Reset / Risk Exposure
	Goal 1:									**Concurrent definition of system requirements and solutions**	
	Critical Success Factor 1.1									**Understanding of stakeholder needs: capabilities, operational concept, key performance parameters, enterprise fit (legacy). Evidence of ability to analyze strengths and shortfalls in current-system operations via:**	4
4	1.1(a)	○	○	○	○	○	○	◉	○	Participatory workshops, surveys, focus groups?	
4	1.1(b)	○	◉	○	○	○	◉	○	○	Operations research techniques: operations data collection and analysis?	
3	1.1(c)	○	○	◉	○	○	◉	○	○	Mission effectiveness modeling and simulation?	
3	1.1(d)	○	◉	○	○	○	○	◉	○	Prototypes, scenarios, stories, personas?	
4	1.1(e)	○	◉	○	○	○	◉	○	○	Ethnographic techniques: Interviews, sampled observations, cognitive task analysis?	

SEPRT and SECRT Concepts of Operation

The SEPRT and SECRT framework and tools provide a way for projects to identify the major sources of program risk due to SE shortfalls. This section summarizes concepts of operation for applying the tools at major milestones, and at other points where SE demonstration shortfalls or other SE EMs such as the INCOSE Leading Indicators have identified likely problem situations and need further understanding of the problem sources and their relative degrees of risk. More detail and examples are provided in the SE EM technical report [36].

The first step in the concept of operations involves collaborative planning by a project's sponsoring decision authority (at a developer level, a program level, or a program oversight level) and its performing organization(s) to reach agreements on the relative priorities of its needed performance aspects and personnel competencies, as measured by the relative program impact of their SEPRT and SECRT question content. The planning necessarily includes consideration of the consistency of these priorities with the project's SE budget, schedule, and contract provisions. This stage frequently identifies inconsistencies between sponsor priorities (e.g., early key performance parameter (KPP) tradeoff analysis) and contract provisions (e.g., progress payments and award fees initially focused on functional specifications and not KPP satisfaction), and enables their timely resolution.

The next step involves evaluation via independent experts of the evidence of adequacy provided by the performers for their ability to perform the desired levels of SE performance within the budgets, schedules, and staffing defined in their SE plans. This should include the rationale for the choices of

evidence preparation such as prototyping, modeling, simulation, benchmarking, exercising, testbed preparation, scenario generation, instrumentation and data analysis; and their associated resource consumption. Subsequently, progress with respect to the plans needs to be monitored; it is best to consider the planned evidence as a first class deliverable, and to include its progress measurement in the project's earned value monitoring system.

Finally, the completed evidence needs to be provided by the performers at each major milestone review, along with the performers' rating of the strength of the evidence and the associated risk level indicated by the SEPRT and SECRT tools. These ratings are then evaluated by independent experts, and adjusted where the evidence is stronger or weaker than the indicated ratings. The revised ratings are discussed and iterated by the sponsors and performers, and used to determine revised SEPRT and SECRT risk levels. These will enable to sponsors and performers to determine the necessary risk mitigation plans, budgets, and schedules for ensuring project success. Again, more detailed scenarios and flowcharts are provided in the SE EM technical report [36].

Summary of Framework and Tool Evaluations

We solicited pilot evaluations of the EM performance and competency frameworks, using the prototype SEPRT and SECRT tools, from industry, government agencies, and academic participants. Because the task re-scoping permitted only a single round of piloting, these initial evaluations were conducted against historical projects and case studies. The tools were successfully piloted against five DoD projects, one NASA project, and one commercial project. They were also analyzed by two industrially-experienced colleagues against detailed case studies of a number of DoD and commercial projects. The application domains piloted included space, medical systems, logistics, and systems-of-systems. Results of the pilot evaluations were reported through a web-based survey tool and detailed follow-up interviews, while the case study evaluations were reported through detailed comments from the reviewers.

Evaluations were generally positive, and the frameworks were found to be useful across all project phases except Production, and against all systems types except "legacy development." The consensus of reviewers was that the frameworks would be most useful in the System Development & Demonstration (SDD) phase, and generally more useful in early phases than later. It was noted, however, that in systems developed using evolutionary strategies, such "early" phases recur throughout the development cycle, extending the usefulness of the frameworks. The evaluations were reported to take 2-5 hours to complete for persons familiar with the projects, with materials that were readily at hand. Also, in reviewing case study material, some evaluators reported that the EM framework was not specific to any particular problem domain (a choice we made to make domain tailoring user-performable via Excel).

Several evaluators reported that the frameworks generated too many high-risk findings, which might make the results too overwhelming to take action. In response to this significant concern, the impact scales were adjusted to make the adjectives better correspond to the quantitative impacts (Critical-Significant-Moderate-Little or No vs. High-Medium-Low-No impact), and a longer risk exposure scale developed to allow more nuanced results.

In addition, the University of Maryland (UMD) Fraunhofer Center (FC) performed preliminary evaluations against the Systemic Analysis Database (SADB), compiled by OUSD (AT&L), and a

mapping between the SEPRT questions and the Defense Acquisition Program Support (DAPS) methodology underlying the SADB results. This evaluation approach allowed analysis of the effectiveness of the frameworks with respect to historical success and failures of the subject projects, and another cross check of the SEPRT coverage. Overall, the coverage mapping indicated that the two were largely consistent, with more domain coverage in the DAPS methodology. A similar mapping was performed between the SECRT and the Defense Acquisition University's SPRDE-SE/PSE Competency Model, with similar results.

Further, a business case analysis for the investment in SE effectiveness evidence was performed, based on data from 161 software-intensive systems projects used to calibrate the COCOMO II cost estimation model and its Architecture and Risk Resolution scale factor. It concluded that the greater the project's size, criticality, and stability are, the greater is the need for validated architecture feasibility evidence (i.e., evidence-based specifications and plans). However, for very small, low-criticality projects with high volatility, the evidence generation efforts would make little difference and would need to be continuously redone, producing a negative return on investment. In such cases, agile methods such as rapid prototyping, Scrum and eXtreme Programming will be more effective. Overall, evidence-based specifications and plans will not guarantee a successful project, but in general will eliminate many of the software delivery overruns and shortfalls experienced on current software projects. Again, more details are provided in the SE EM technical report [36].

Conclusions

DoD programs need effective systems engineering (SE) to succeed.

DoD program managers need early warning of any risks to achieving effective SE.

This SERC project has synthesized the best analyses of DoD SE effectiveness risk sources into a lean framework and toolset for early identification of SE-related program risks.

Three important points need to be made about these risks.

- They are generally not indicators of "bad SE." Although SE can be done badly, more often the risks are consequences of inadequate program funding (SE is the first victim of an underbudgeted program), of misguided contract provisions (when a program manager is faced with the choice between allocating limited SE resources toward producing contract-incentivized functional specifications vs. addressing key performance parameter risks, the path of least resistance is to obey the contract), or of management temptations to show early progress on the easy parts while deferring the hard parts till later.

- Analyses have shown that unaddressed risk generally leads to serious budget and schedule overruns.

- Risks are not necessarily bad. If an early capability is needed, and the risky solution has been shown to be superior to the alternatives, accepting and focusing on mitigating the risk is generally better than waiting for a better alternative to show up.

The results of the SEPRT and SECRT pilot assessments, the DAPS and SADB comparative analysis, and the quantitative business case analysis for the use of the SE EM framework, tools, and operational

concepts is sufficiently positive to conclude that implementation of the approach is worth pursuing. Presentations at recent workshops have generated considerable interest in refining, using, and extending the capabilities and in co-funding the followon research. However, the framework and prototype tools have been shown to be largely efficacious only to date for pilot projects done by familiar experts in a relatively short time. It remains to demonstrate how well the framework and tools will perform on in-process MDAPs with multiple missions, performers, and independent expert assessors.

Some implications of defining feasibility evidence as a "first class" project deliverable are that it needs to be planned (with resources), and made part of the project's earned value management system. Any shortfalls in evidence are sources of uncertainty and risk, and should be covered by risk management plans. The main contributions of the SERC SE EM project have been to provide experience-based approaches and operational concepts for the use of evidence criteria, evidence-generation procedures, and SE effectiveness measures for monitoring evidence generation, which support the ability to perform evidence-based SE on DoD MDAPs. And finally, evidence-based specifications and plans such as those provided by the SERC SE EM capabilities and the Feasibility Evidence Description can and should be added to traditional milestone reviews.

As a bottom line, the SERC SE capabilities have strong potential for transforming the largely unmeasured DoD SE activity content on current MDAPs and other projects into an evidence-based measurement and management approach for both improving the outcomes of current projects, and for developing a knowledge base that can serve as a basis for continuing DoD SE effectiveness improvement.

References

[1] Air Force. Probability of Program Success Operations Guide. https://acc.dau.mil/GetAttachment.aspx?id=19273&pname=file&aid=976&lang=en-US

[2] Al Said, M. 2003. Detecting Model Clashes During Software Systems Development. Doctoral Thesis. Department of Computer Science, University of Southern California.

[3] ASN (RD&A). Guidebook for Acquisition of Naval Software Intensive Systems. September 2008. http://acquisition.navy.mil/organizations/dasns/rda_cheng.

[4] Basili, V., Gianluigi, C., Rombach, D. 1994. The Experience Factory. In Encyclopedia of Software Engineering. John Wiley & Sons, 469-476.

[5] Bass, L., Clements, P., Kazman, R., Klein, M. Models for Evaluating and Improving Architecture Competence. CMU/SEI-2008-TR-006 (April 2008).

[6] Beck, K. 1999. Extreme Programming Explained. Addison-Wesley.

[7] Boehm, B. 1996. Anchoring the Software Process. IEEE Software (Jul. 1996), 73-82.

[8] Boehm, B., et al. 2000. Software Cost Estimation with COCOMO II. Prentice Hall.

[9] Boehm, B., Ingold, D., Madachy, R. 2008. The Macro Risk Model: An Early Warning Tool for Software-Intensive Systems Projects http://csse.usc.edu/csse/event/2009/UARC/material/Macro%20Risk%20Model.pdf

[10] Boehm, B. and Lane, J. 2009. Incremental Commitment Model Guide, version 0.5.

[11] Boehm, B. and Lane, J. 2007. Using the ICM to Integrate System Acquisition, Systems Engineering, and Software Engineering. CrossTalk (Oct. 2007), 4-9.

[12] Boehm, B., Port, D. and Al Said, M. 2000. Avoiding the Software Model-Clash Spiderweb. IEEE Computer (Nov. 2000), 120-122.

[13] Boehm, B., Valerdi, R., and Honour, E. 2008. The ROI of Systems Engineering: Some Quantitative Results for Software-Intensive Systems. Systems Engineering, Fall 2008, pp. 221-234.

[14] Componation, P., Youngblood, A., Utley, D., Farrington. Assessing the Relationships Between Project Success, and System Engineering Processes. http://csse.usc.edu/csse/event/2009/UARC/material/Project%20Success%20and%20SE%20Processes%20-%20UAHuntsville%20-%20Componation.pdf

[15] Elm, J., Goldenson, D., Emam, K., Donatelli, N., Neisa, A., NDIA SE Effectiveness Committee. A Survey of Systems Engineering Effectiveness – Initial Results. CMU/SEI-2008-SR-034 (Dec. 2008), http://www.sei.cmu.edu/reports/08sr034.pdf

[16] Glass, R. 1998. Software Runaways. Prentice Hall.

[17] INCOSE. 2007. INCOSE Systems Engineering Handbook v. 3.1. INCOSE-TP-2003-002-03.1.

[18] Kruchten, P. 1999. The Rational Unified Process. Addison Wesley.

[19] Maranzano, J., et al. 2005. Architecture Reviews: Practice and Experience. IEEE Software (Mar./Apr. 2005).

[20] National Research Council (U.S.). 2008. Pre-Milestone A and Early-Phase System Engineering: A Retrospective Review and Benefits for Future Air Force Systems Acquisition, Washington D.C.: National Academies Press.

[21] Office of the Deputy Under Secretary of Defense for Acquisition and Technology, Systems and Software Engineering, Defense Acquisition Program Support (DAPS) Methodology, Version 2.0 (Change 3), March 20, 2009.

[22] Office of the Director of National Intelligence (ODNI), Subdirectory Data Collection Tool: Systems Engineering.

[23] Pew, R. and Mavor, A. 2007. Human-System Integration in the System Development Process: A New Look. National Academy Press.

[24] Roedler, G., Rhodes, D. Systems Engineering Leading Indicators Guide. http://csse.usc.edu/csse/event/2009/UARC/material/SELeadingIndicators2007-0618.pdf

[25] Royce, W. 1998. Software Project Management. Addison Wesley.

[26] Royce, W., Bittner, K., and Perrow, M. 2009. The Economics of Iterative Software Development. Addison Wesley.

[27] Schwaber, K. and Beedle, M. 2002. Agile Software Development with Scrum. Prentice Hall.

[28] SISAIG. 2004. System-Software Milestone/Review Decision Support Framework. US/UK/AUS Software Intensive System Acquisition Improvement Group.

[29] Standish Group 2009. CHAOS Summary 2009. http://standishgroup.com.

[30] United States Government Accountability Office. Defense Acquisitions Assessments of Selected Weapon Programs. http://www.gao.gov/new.items/d09326sp.pdf

[31] USC. Anchor Point Feasibility Evidence. http://csse.usc.edu/csse/event/2009/UARC/material/MBASE_Guidelines_v2.4.2%20FRD.pdf

[32] USC. 2007. Evaluation and Calibration of Macro Risk Model (August 2007).

[33] USC. Macro Risk Tool. http://csse.usc.edu/csse/event/2009/UARC/material/Macro%20Risk%20Model%20v%2012.1.xls

[34] Weitekamp, M., Verma, D. Leading Indicators of Program Success and Failure. Stevens Institute http://csse.usc.edu/csse/event/2009/UARC/material/Stevens%20Leading%20Indicators%20Project.pdf

[35] Williams, C., Derro, M. NASA Systems Engineering Behavior Study. NASA Office of the Chief Engineer (Oct. 2008).

[36] TBD: put in full list of authors (including yourself), final tech report title, 30 Sept 2009 date, and TBD reference to a SERC web site location.