28 August 2007

Final Report

System Safety Performance Level Model

Doc. No. S-07-00400

Prepared for:
Concurrent Technologies Corporation
1225 S. Clark Street, Suite 500
Arlington, VA 22202

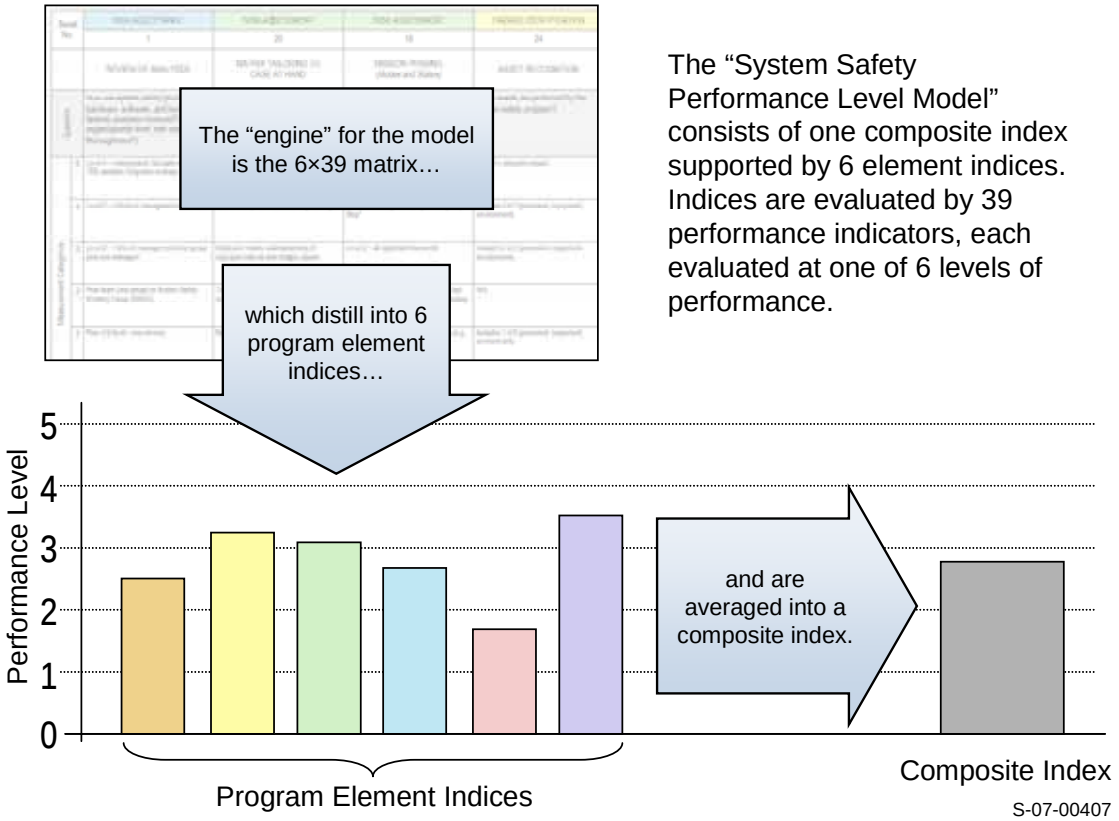
Prepared by:
APT Research, Inc.
4950 Research Drive
Huntsville, Alabama USA 35805

EXECUTIVE SUMMARY

The Defense Safety Oversight Council (DSOC) Acquisition and Technology Programs (ATP) Task Force funded development of the System Safety Performance Level Model. The model serves as a useful tool to gauge the health of a safety program at any stage of the lifecycle of the program. Experience has proven that a strong safety program results in significant savings to the program, reduced need for late application of corrective retrofits, and often more effective systems at lower overall cost. Just as early discovery and correction of safety program weaknesses favors cost savings, so does early opportunity for an effective safety program to influence design favor more effective system design outcomes at reduced overall cost.

An approach for evaluating system safety program effectiveness was the outcome of a 2006 workshop devoted exclusively to the topic. System safety practioners in attendance responded to data collection questions probing performance within key elements of practice. One hundred fifty-five (155) performance indicators for a good safety program were drafted. This work narrowed the initial 155 indicators into a condensed set of 39 factors that became the basis of the System Safety Performance Level Model. Full design criteria are specified here as well as detailed guidance for applying the model.

The System Safety Performance Level Model depicted below consists of a recommended performance scale (0-5), 39 inquiry items, detailed data collection sheets, and a means to track the data.



Advice from a panel of experienced practitioners was sought for use to guide optimum timing for data collection among program lifecycle phases. The expressed preference is to administer the model early, and repeatedly thereafter. In a beta test of the model, 18 practitioner interviews were conducted and the results analyzed. A second workshop was held to evaluate the model and examine the results and findings of the beta test. Panel participants arrived at consensus on the overall method and requested the addition of special emphasis in the areas of software safety and human factors.

Chief advantages of the method are efficacy in use and immediacy of feedback on program performance. The use of this model will rapidly provide program assessment data which heretofore have only been available via extensive program review and tedious after-the-fact analyses.

Contents

Appendices

Figures

Tables

Acronyms

AMCOM	Aviation and Missile Command
APT	APT Research, Incorporated
ATP	Acquisition and Technology Programs
BS	Bachelor of Science
CCA	Cause-Consequence Analysis
CIH	Certified Industrial Hygienist
COTS	Commercial-off-the-Shelf
CPI	Critical Program Information
CSP	Certified Safety Professional
DoD	Department of Defense
DODI	Department of Defense Instruction
DSOC	Defense Safety Oversight Council
ESOH	Environmental, Safety, and Occupational Health
ETA	Event Tree Analysis
FAA	Federal Aviation Administration
FHA	Failure Hazard Analysis
FMEA	Failure Modes and Effects Analysis
GEIA	Government Electronics and Information Technology Association
GOTS	Government-off-the-Shelf
HAZOP	Hazardous Operation
HIS	Human Systems Integration
IPT	Integrated Product Team
MIL STD	Military Standard
NASA	National Aeronautics and Space Administration
NDI	Non-Developmental Items
PE	Professional Engineer
PHA	Preliminary Hazard Analysis
SDD	System Development and Demonstration
SEAC	Safety Engineering and Analysis Center
SSPP	System Safety Program Plan
SSWG	System Safety Working Group
U.S.	United States

INTRODUCTION

The Defense Safety Oversight Council (DSOC) Acquisition and Technology Programs (ATP) Task Force funded the development of the System Safety Performance Level Model. The DSOC mission is to investigate and recommend or implement changes to policies, procedures, initiatives, education and training, and investments to ensure that acquisition programs address safety throughout program lifecycle. The model serves as a useful tool to gauge the health of a safety program throughout its lifecycle.

This report first describes the model resulting from this effort (Section I) and then describes the development process (Section II).

SECTION I. THE SYSTEM SAFETY PERFORMANCE LEVEL MODEL

The System Safety Performance Level Model consists of a recommended performance scale, 39 inquiry items, detailed data collection sheets, and a means to track the data. The complete model summarized in Figure 1 is at Appendix I.

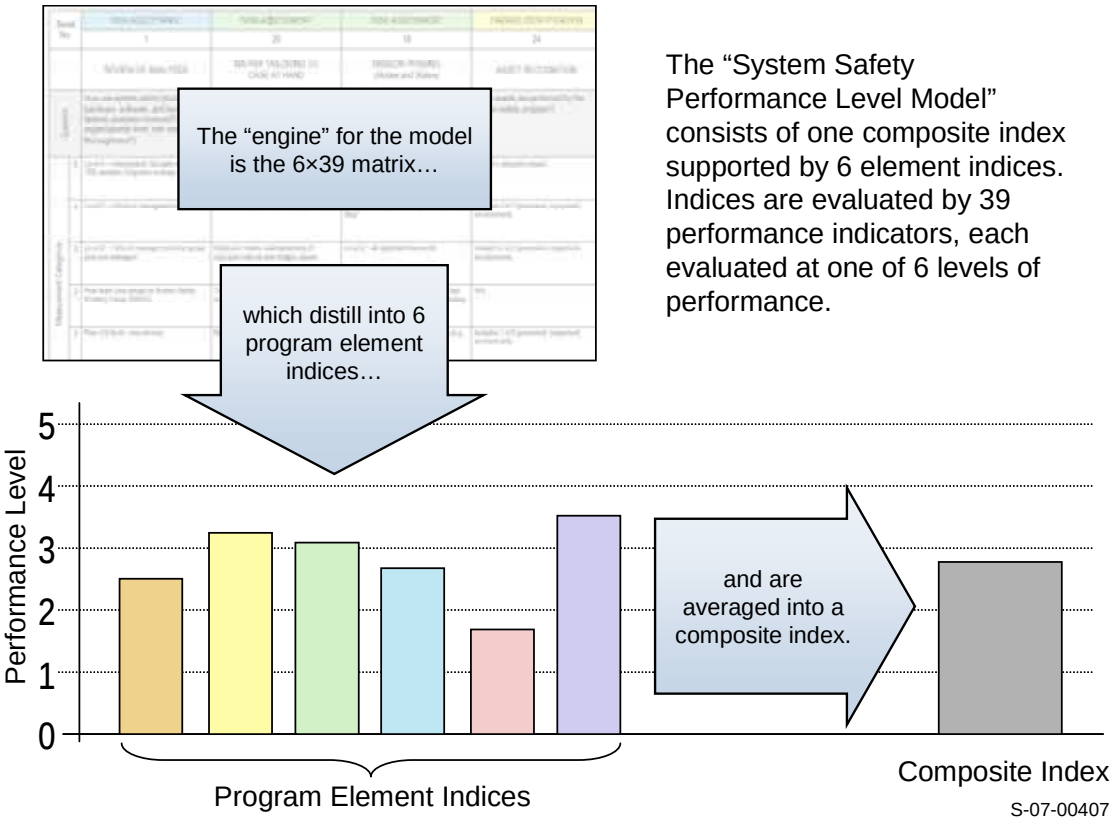


Figure 1. *The System Safety Performance Level Model*

THE MODEL DESIGN CONCEPT

There are several important considerations in designing an evaluation model to represent the overall state of a safety program. The design goals were as follows:

1. Include major elements of a system safety program
2. Provide ease of administration
3. Have the minimum essential inquiries to determine program performance.

The design includes six evaluation elements that correspond to the six major elements of a safety program (Figure 2). Each element is assessed using a number of inquiry items. Each inquiry is answered (rated) by selecting one of six performance levels.

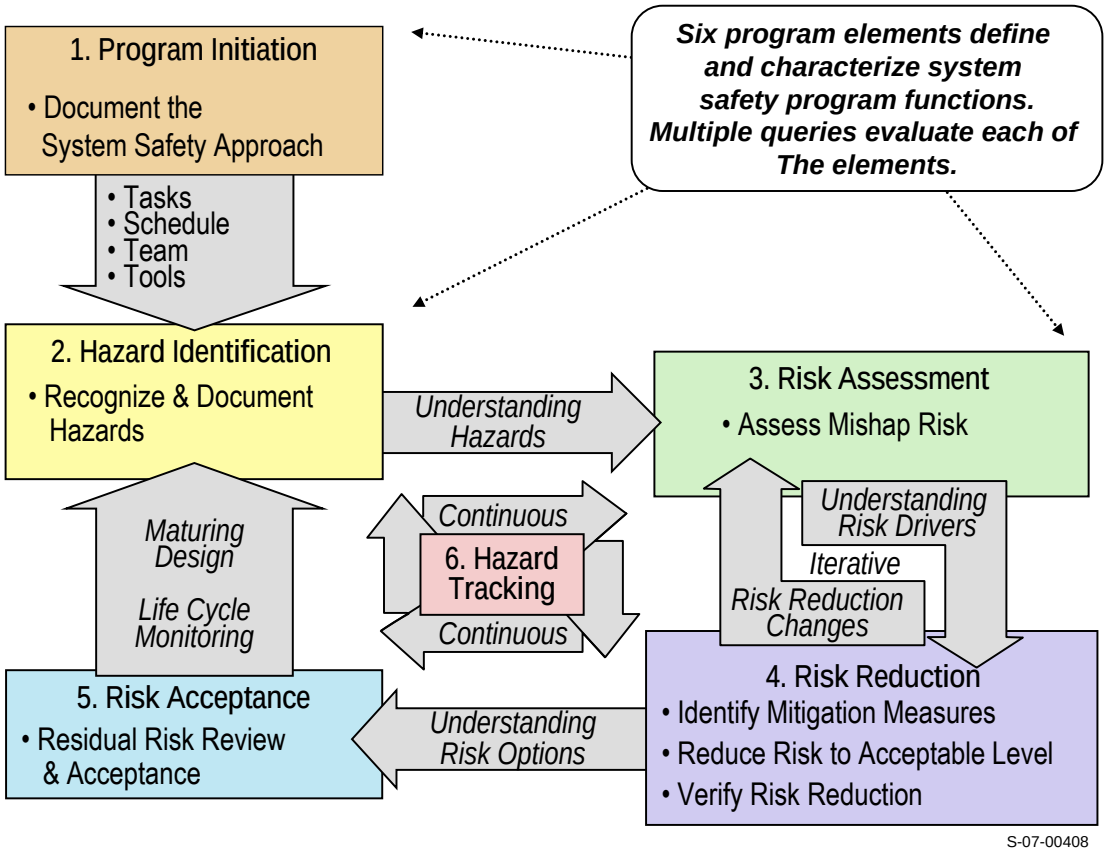


Figure 2. *Safety Program Elements*

In selecting the total number of queries, significant discussion focused on how many would be needed.

Anecdotal evidence from other well known indices illustrated that, if selected carefully, an index with as few as 20 queries will give as concise a measurement as one with 500queries. Shown in Figure 3 is the indistinguishable overall measurement of the U.S. Stock Market performance as measured by the DOW (with 19 queries) and the S&P (with 500 queries). This logic supported the sizing of this model.

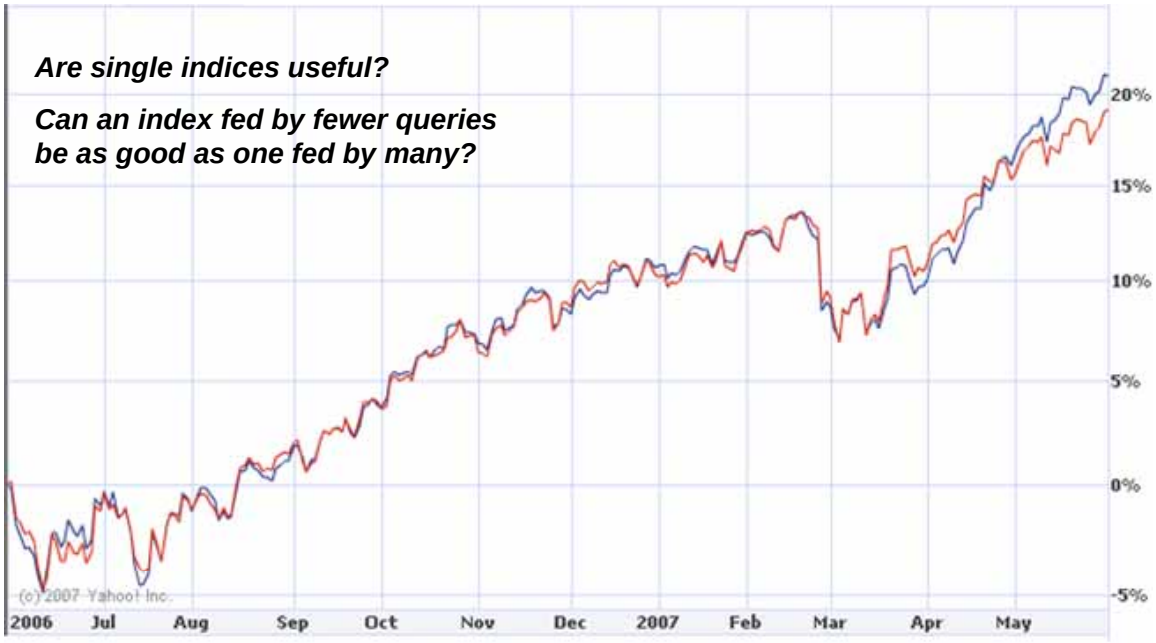


Figure 3. *The Use of an "Index Fed by Multiple Subfactors"*

POTENTIAL APPLICATIONS OF THE MODEL

Potential applications of the System Safety Performance Level model include gauging the performance adequacy of a system safety program, in any phase of its life cycle, separately identifying areas of specific performance strengths and weaknesses. The model can also be used in evaluating the adequacy of program performance with a field-developed standard (when normed through multiple field applications).

It is expected that such a measurement model would be applicable during acquisition as a tool to be used in ensuring appropriately effective system safety program capability, and thereafter for monitoring program effectiveness and for guiding improvement when necessary.

A beta test performed as a separate task established that the model can point to needed improvements in a safety program, and that the model has good potential as a management tool.

Mishap Reduction

The ability to manage the risks of hazards more effectively and with reduced turn-around leads to an improved capability to reduce both the number and the severity of mishaps. While it is not foreseeable that the model will accurately predict the number of accidents that will be reduced, that is not its purpose. Instead, it will ensure an improved capability to produce that reduction.

Human error is one of the greatest causes of accidents. Human error is often mitigated though the use of procedures and training. Far more effective mitigation means are preferable. If a system safety program is evaluated early in the program a manager can predict whether more effective mitigation measures will be adopted.

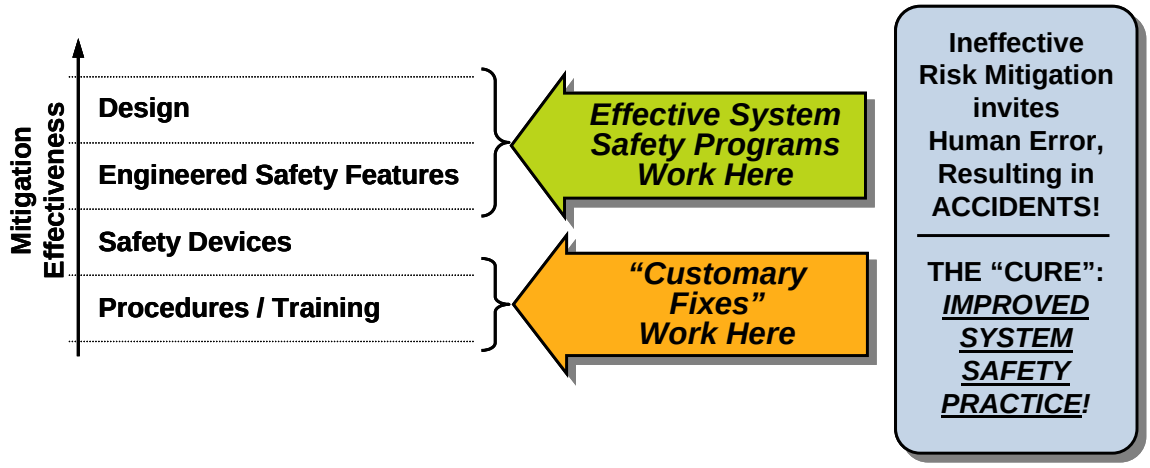


Figure 4. *System Safety Drives Effective Risk Mitigation*

Implementation

System safety is a subset of systems engineering (SE); those that have funded this effort through the ATP Task Force have responsibility for SE across DoD. This may be a good starting place to implement a DoD tool across the services.

The System Safety Performance Level Model could readily be made available to those desiring to use it by posting on web sites such as the OSD ATP Task Force safety initiative web site, the AT&L Knowledge Sharing System/Defense Acquisition University, or others.

DATA COLLECTION OPTIONS

There are at least six approaches available for use in collecting data to evaluate a safety program. The approach selection will vary with the circumstances of the need.

1. A Safety Program Guidelines/Checklist approach is useful in policy documents or manuals. The model readily applies to this approach.
2. A self-evaluation approach could be useful as an in-house improvement application. The model readily applies to this approach.
3. A questionnaire option would be inexpensive. However, this application could be biased by self-interest. The model readily applies to this approach.
4. An interview option would be in expensive and assures some norming against bias. The model readily applies to this approach. This approach was selected for the beta test of the model.
5. An in-depth survey may include an additional inspection by headquarters or the sponsoring government agency. The model could be utilized as a part of this approach.
6. A complete audit would include a thorough investigation of the safety program, supported by the model.

ADVANTAGES AND WEAKNESSES OF THE MODEL

Advantages of the Model

The advantages of using the System Safety Performance Level Model include:

- Examining and expressing the efficacy of a system safety program while it's "up and running," with only modest man-hour interruption of normal activities.
- Identifying performance inadequacies and providing targeted feedback that can direct positive corrective guidance without wasted effort.
- Shortening the cycle separating flawed practice and its discovery, thus shortening the "temporal distance" that often cripples leading indicators.
- Providing fast results — conserving calendar time and man hours, hence cost, by comparison with in-depth reviews of final analytical products.
- Excellent potential as a management tool. Management can see program strengths and/or weaknesses with swift turn-around.
- Improves management of risks and hazards.
- No special expertise required to administer beyond familiarity with the system safety practice.
- The model is equally applicable by Government or Contractor organizations, resulting in improved guidance on an internal "no-fault" basis.
- Return on Investment: This project improves the overall acquisition process by integrating evaluation measures that would: give Government organizations a means to specify or evaluate system safety programs; give mature industry and Government programs a means to "certify" existing maturity; and give immature industry or Government programs a "way ahead" toward greater maturity. This capability provides a common language supporting effective execution for system safety at each phase in the life cycle effort for acquisition programs.

Potential Weakness of the Model

A potential weakness of the System Level Performance Level Model is that, as an inferential method, it examines professional practitioner "viewpoints" on practices and their adequacy rather than evaluating the caliber of the final analytical product. Workshop participants were willing to make this sacrifice in "measurement certainty" in return for quickly obtained results.

SECTION II. BACKGROUND OF MODEL DEVELOPMENT

THE NEED

A need has long been recognized for a means to determine the "goodness" of applied system safety practice. Just how effective is a particular system safety program at identifying and controlling the threats posed by hazards within a system? The chief deterrent to success in

measuring the efficacy of system safety practice is markedly similar to an obstacle found in evaluating methods for disease prevention and control. In both system safety practice and disease control, the purpose is to manage the threat of adverse outcomes. In both fields, a direct measure of the degree to which unwanted consequences are averted can only be gauged empirically. Experiments must be carried out in which the performance of analyzed systems — or of treated patients — is compared to the performance of otherwise identical untreated ones. In each field, reducing the probability and severity of adverse outcomes is both the desired result and a measure of success.

In system safety practice, where reducing the probability of loss events is the goal, such a direct empirical approach would be exorbitantly costly. Systems identical in purpose must be built, differing only in that one will have had the benefit of system safety application in its design, construction and operation, and the other will not. In immunology it is customary that a population of patients will be available to participate in controlled clinical trials to determine the differences in outcomes among cases that are treated and those that are not. Such a direct approach is clearly impracticable in system safety practice. An alternative method is needed.

SATISFYING THE NEED

A Workshop Approach

A workshop to address the need and to devise a way of satisfying it was hosted on 21-22 February 2006, by the Safety Engineering and Analysis Center (SEAC) in Huntsville, Alabama. Forty-six attendees participated in the two-day workshop. The participants, listed in Appendix II, included system safety practitioners from the United States (U.S.) Department of Defense (DoD), the Federal Aviation Administration (FAA), the National Aeronautics and Space Administration (NASA), Great Britain, Sweden, Canada, industry, government, and academia. Many were members of the G-48 Committee of the Government Electronics and Information Technology Association (GEIA), which had met at the same location during the same week. (It is the G-48 Committee that is charged with recommending improvements in the leading standard governing system safety practice, MIL-STD-882.) To address the need for evaluating the quality of practice prevalent in a system safety program, the 46 attendees participating in the workshop developed a large family of performance-related queries. This output provided the 155 queries that formed the basis for this effort. These were questions of kinds that would explore the effectiveness of a system safety program's pertinent.

MODEL ORIGINS AND DESIGN

Following the workshop, a panel of SEAC safety professionals developed and evaluated a variety of candidate approaches for using the queries developed by the workshop attendees to the gauging of system safety program efficacy. Chief among them were:

- Detailed reviews of analyses produced by the program undergoing evaluation. This candidate was rejected for two reasons - the cost in man hours and the delay in acquiring results.
- Comparisons of the characteristics and attributes of the program being evaluated with characteristics and attributes of other programs of known quality. This candidate was

rejected because of the difficulty of reaching accord on the quality of performance of proposed comparison programs.

- An evaluation of the efficacy of system safety program practice based on analyzing performance-related information acquired through interviews.

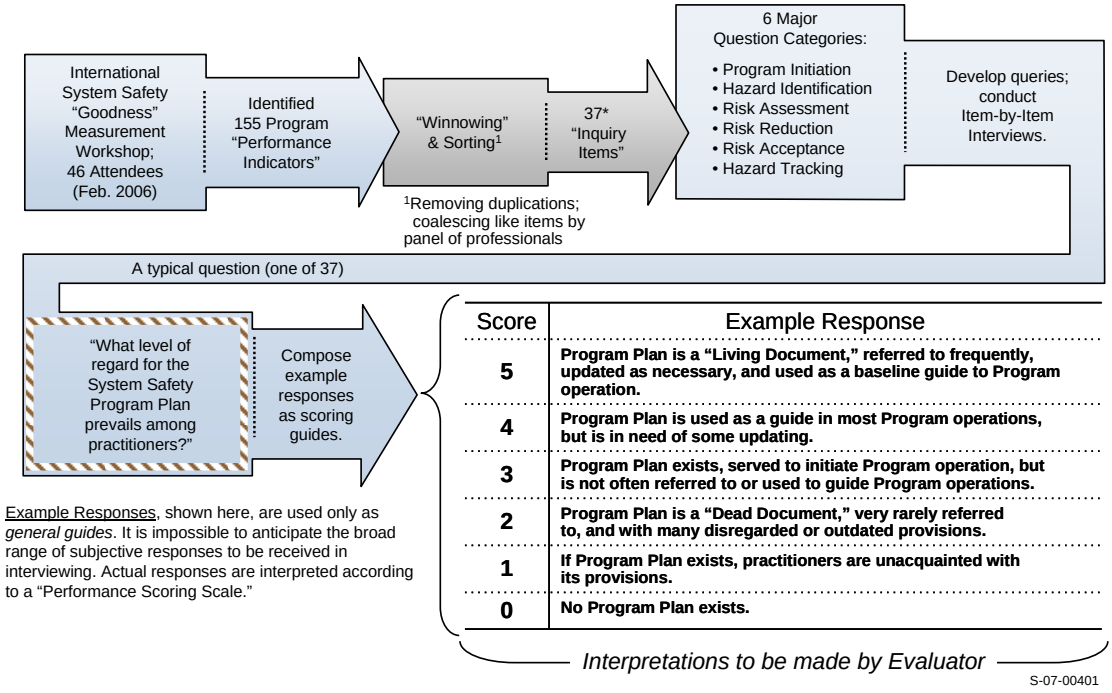
For this effort an approach based on evaluating the efficacy of system safety program practice by analyzing performance-related information acquired through interviews among hands-on practitioners working within the program being examined was adopted. Possible barriers to the success of this approach were foreseen:

- Interviewees would not be likely to report on adverse aspects of program performance because of fear of reprisals. It was reasoned that these fears could be allayed by providing interviewees absolute assurances of privacy.
- Because the method is inferential rather than direct, the results would be technically "impure." The alternative, direct program evaluation by detailed engineering reviews of analytical program products, was viewed as impracticable for two reasons: because such an approach is time consuming, it would defeat the purpose of providing prompt feedback of results from which corrective guidance might be made available to program management; moreover, the ready availability of adequately qualified reviewers was thought to be in doubt.

To overcome these two potential shortcomings was accepted as a challenge in developing the interview questions and in devising an effective interview style.

The participants in the February 2006 workshop had developed performance-related queries of kinds that would explore the effectiveness of a system safety program's pertinent features. The overall flow of development activity is shown in Figure 5. As shown there, the workshop participants had developed 155 "performance indicators." Following the workshop, these performance indicators were reduced to 37 "inquiry items" by a panel of four experienced SEAC system safety practitioners. The purpose of this reduction was:

- to eliminate unintended duplications
- to coalesce closely related items
- to distribute the items among the six major functional elements of system safety program operation, as shown in Figure 6, which also indicates the number of items devoted to exploring each of the six major program elements. (Note that, in the judgment of the panel members, several of the six elements can be adequately probed with only a few inquiry items while others require many.)
- to produce a battery of topics that could be developed as interview questions.



Example Responses, shown here, are used only as *general guides*. It is impossible to anticipate the broad range of subjective responses to be received in interviewing. Actual responses are interpreted according to a "Performance Scoring Scale."

Figure 5. *Developing the Evaluation Tool*

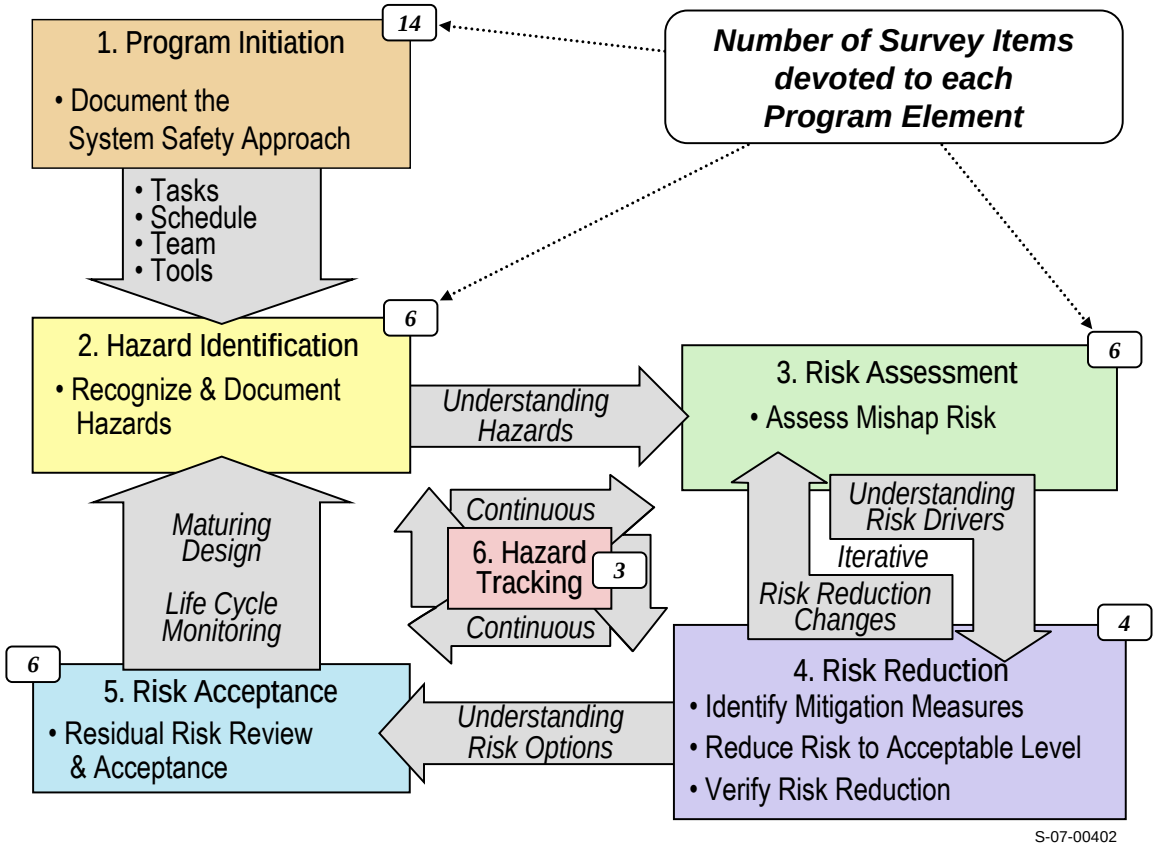


Figure 6. *Six System Safety Program Elements*

Figure 7 illustrates the method used in removing unintended duplications and in coalescing like items.

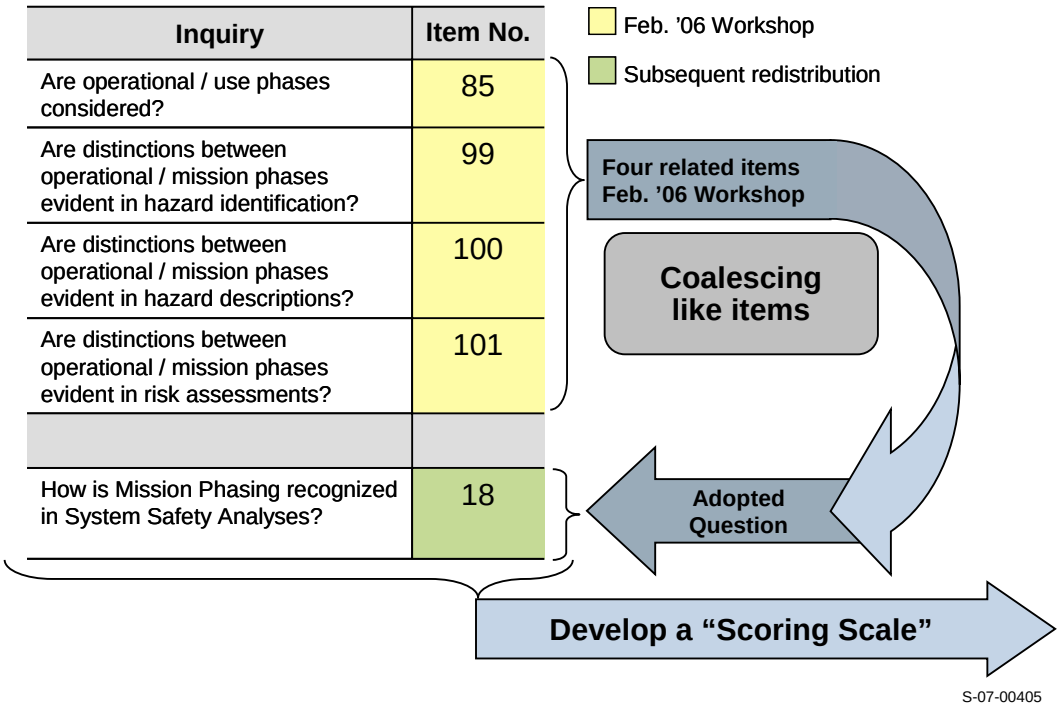


Figure 7. *Developing Questions for the Model*

Ten design criteria were established to guide model development. These were based largely on guidance drawn from the abundant literature available on the subject (Refs.1, 2, and 3). The criteria appear as Table 1. It was recognized as important to formulate the set of items to adequately probe system safety program performance without consuming an inordinate amount of interviewee time in collecting responses (see Design Criteria 4 and 8, Table 1.)

Table 1. Data Gathering Design Criteria

1	Interview ≥ 10% of program practitioners (Increase sample size if data scatter indicates need – See Refs. 2 and 3 for guidance on selecting and adjusting sample size.)
2	Employ one-on-one interviews (one interviewer, one interviewee)
3	Protect Interviewee privacy (confidentiality must be assured)
4	Limit interviews to minimum number needed for evaluation—data scatter governs
5	Structure questions for discussion-style, rather than one-word response
6	Explore both "traditional" and "cutting-edge" system safety practices
7	Evaluate responses according to multiple "performance determinants" developed and sanctioned by well-recognized practitioners
8	Interview duration ≈ one hour each
9	Provide adequate resolution of results (i.e., ≥ 5 discrete measurement levels)
10	Provide results interpretable as guidelines for System Safety Program improvement (if needed)

As shown in Figure 5, the effort then moved to devising, for each of the inquiry items, a question that could be posed to each member of a sample population of practitioners within a system

safety program undergoing evaluation. Responses to these questions are then to be scored against a common scale to gauge the "goodness" of the relevant element of practice. For each of the questions, a set of example responses was composed at staged levels of quality. These range from the extreme of absent practice to an opposite superlative. A typical question and example responses appear in Figure 5. The example responses are intended only as illustrative guides to be used in conjunction with a "Performance Scale," constructed to support interpretation of interviewee responses and to arrive at a score for each response. The Performance Scale is shown here as Table 2.

Table 2. Performance Scale

Score	Descriptor
5	Excellent: Superlative in both concept and performance
4	Good: Satisfying all needs, in letter and in spirit
3	Fair: Satisfying most needs, but with one or a few notable flaws
2	Marginal: Failing to satisfy one or more key needs
1	Minimal: Present, but having little or no value to the Program
0	Null or Absent: Virtually non-existent/not practiced

Use of such a performance scale in combination with the example responses is necessary because it is impossible to anticipate the very broad range of subjective responses to be encountered during interviews. The scoring system resulting from this approach is made intentionally demanding. Unless this is done, performance distinctions between outstanding programs cannot not be gauged. For many items, all would score "off-scale high" by unmeasured amounts.

ADMINISTERING THE MODEL

Inquiry Items

The questions making up the central body of the interviews appear in Table 3. Also shown there is the system safety program element with which each question is associated. The questions are not listed in the sequence in which they are to be presented to interviewees. During the interviews, no two consecutive questions are to address the same program element. This practice follows advice found in the references on conducting interviews for the purpose of data gathering. The purpose is to discourage adjacent answers from biasing one another.

Table 3. Inquiry Items

#	Question	Program Element*
1	How are system safety (including hardware, software, and human factors) analyses reviewed? (At what organizational level; with what thoroughness?)	5
2	How are waivers, exceptions, and other non-compliances managed by the system safety program?	5
3	How does system safety (including hardware, software, and human factors) manning allocation compare to actual needs?	1

#	Question	Program Element*
4	Are the six program elements and functions recognized? (i.e., program initiation, hazard identification, risk reduction, risk acceptance, hazard tracking)	1
5	What authority does the system safety organization have?	1
6	Which model best describes system safety organization structure and reporting level?	1
7	Is management practice consistent with current standards?	1
8	What level of regard prevails for the system safety program description documents (including hardware, software, and human factors)?	1
9	Does the system safety program specifically address Environmental, Safety, and Occupational Health (ESOH) hazards?	1
10	How is system safety practice linked to the "illities" – e.g., Reliability, Availability, Maintainability?	6
11	How often, long-term average, has the program made system safety organizational/personnel changes?	1
12	How is acceptance of residual risk documented/reported?	5
13	Are well-trained system safety personnel involved in the specification and design process? Do the design specifications reflect the impact of safety inputs?	1
14	How are hazard analyses results tracked, recorded, documented, updated, and closed?	3
15	How is risk assessed for a single hazard that threatens multiple protected assets? Use clarifying example.	3
16	How are risk tolerance limits selected for programs?	5
17	What system safety risk summation practices are employed?	5
18	How is mission phasing (modes and states) recognized in system safety (including hardware, software, and human factors) analysis?	3
19	How is exposure interval selection made?	3
20	How is risk assessment matrix tailoring done?	3
21	In assessing risk, how is uncertainty addressed or characterized?	4
22	How is the hierarchy of mitigation precedence treated?	4
23	Are risk assessments quantified?	4
24	What assets are protected by the system safety program?	2
25	How are hazard tracking and safety requirement traceability implemented for system safety (including hardware, software, and human factors)?	6
26	How are hazards identified?	2
27	What hazard inventory-type analytical techniques are used?	2
28	What logic tree analytical techniques are used?	2
29	Are safety hazard data maintained up-to-date?	6
30	How are hazards understood by those working in the program?	2
31	Is analytical tool selection tailored to the needs of individual system peculiarities and needs?	3
32	Are newly discovered hazards reported and mitigation plans formulated promptly?	4
33	What level of training in system safety (including hardware, software, and human factors) have you achieved?	1

#	Question	Program Element*
34	How many years of direct, full-time equivalent experience have you had in system safety (including hardware, software, or human factors) practice?	1
35	What applicable credentials do you hold?	1
36	How would you characterize the system-specific knowledge of system safety staff members?	1
37	How would you characterize organizational safety culture?	1
38	For software-intensive systems, how are the software personnel involved as contributors in identifying the hazards and mitigators?	2
39	How is the analysis of Government-off-the-Shelf (GOTS)/Non-Developmental-Items (NDI)/Commercial-off-the-Shelf (COTS)/software reuse addressed in the system safety program (including hardware, software, and human factors)?	3

* 1 Program Initiation	2 Hazard Identification	3 Risk Assessment	4 Risk Reduction	5 Risk Acceptance	6 Hazard Tracking

Following the structured question and answer portion of each interview, the interviewer is to open the session to an informal discussion. In this period, the interviewee is to be encouraged to express his overall impressions of the system safety program. The interviewer takes notes on pertinent information offered in the discussion. Although this portion of the interview is not to be scored, the discussion will sometimes result in a request from the interviewee to alter the score assigned to one of the questions making up the formal portion of the interview. These requests should be honored.

CONDUCTING THE INTERVIEWS

The manner in which the interviews are conducted is of great importance to the success of the method. An Interviewer's Guide (Appendix III) has been prepared as an aid to this process. This guide draws heavily upon information to be found in the references and on practical lessons learned in administering the interviews in a beta test of the method, discussed later. The guide presumes no prior experience in carrying out such a function by the individual serving as interviewer.

DATA REDUCTION

Average of Averages

Data collected in the interviews are reduced to produce average scores. Scores for all questions are given equal weight in determining program element scores, and the program elements are weighted equally in determining overall program score. (Should it be desired, both question scores and program element scores can be given differing individual weights.)

Data reductions are carried out using formulas of the form:

$$S_i = \frac{\sum_{\ell=0}^{\ell=5} (S_\ell \times n_\ell) W_i}{P_i}$$

Here,

S_i = score for i^{th} item

ℓ = performance level (0, 1, 2 ··· 5)

S_ℓ = i^{th} item score for ℓ^{th} performance level

n_ℓ = number of interviewees voting at performance level ℓ for item i.

P_i = Population voting for the i^{th} item

W_i = Importance weight assigned to item i. (Throughout this program evaluation, $W_i = 1$.)

This "average of averages" data reduction approach is applied at all data levels, e.g.:

- Averaging item responses within program elements
- Averaging program element scores to produce the overall score

Should an interviewee decline or be unable to respond to an interview question, the absence of a data entry results in an adjusted value of the population term, P_i. Thus the term S_i represents the true, average, per-item score of "votes" actually cast.

A data recording and reduction software routine has been prepared for use in accumulating and displaying results. It is provided in Microsoft Excel™ format, separately from this report.

EXAMPLE RESULTS

A data collection example and results that it portrays are illustrated in Figure 8 and Figure 9:

- Figure 8: Here, we see a reproduction of one page of a data collection guidance packet. The questions shown are posed in the left-to-right order indicated. Responses are recorded by interviewer checkmarks made directly on hard copy pages of the packet for later entry into the computer database.
- Figure 9: The results of the interviews appear as a bar chart, as plotted by the Excel™ spreadsheet routine described above.

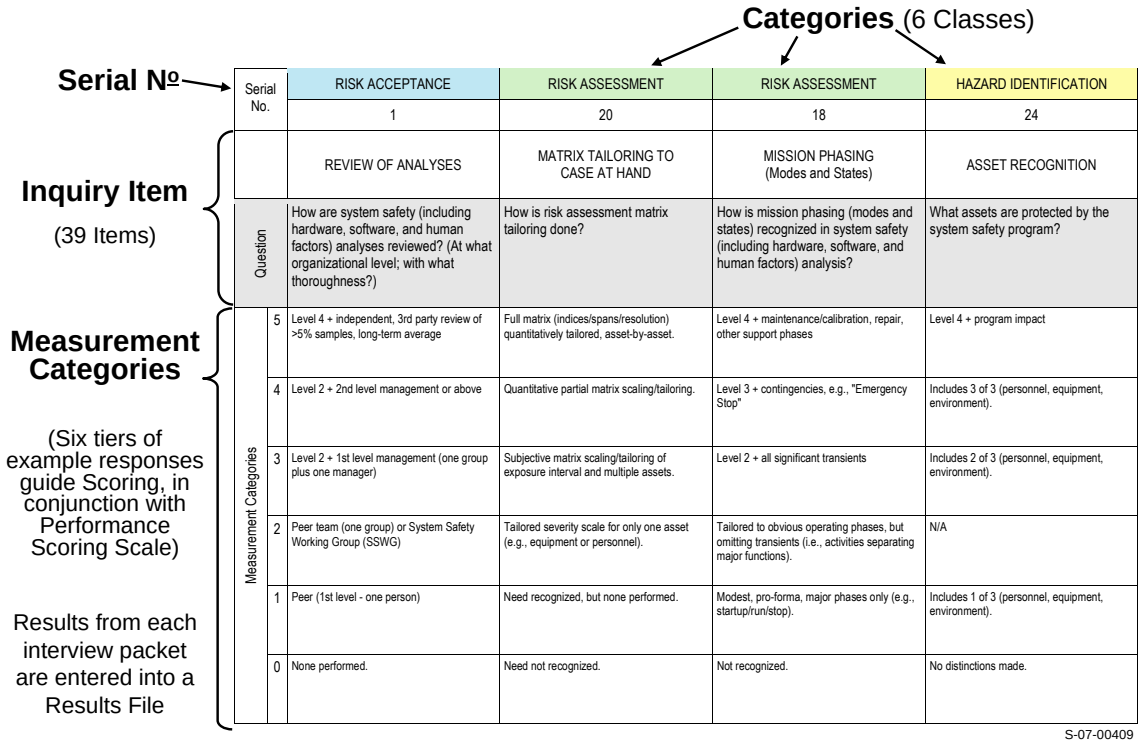


Figure 8. *Data Collection Sheet (example)*

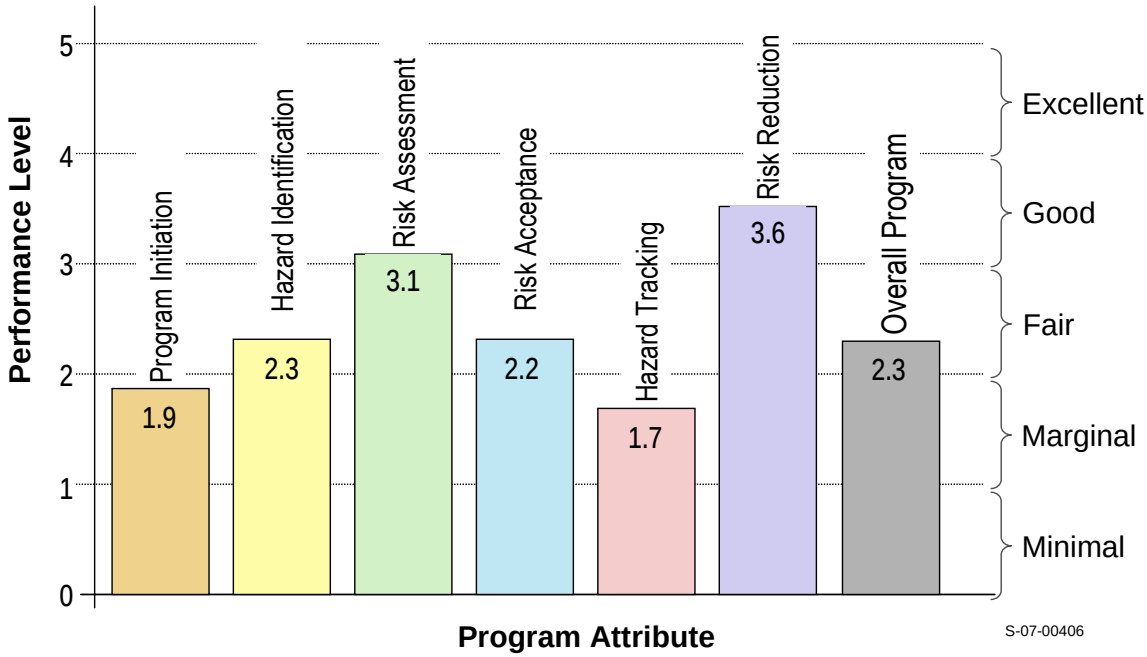


Figure 9. *Sample Results*

Interpretation of the results is self explanatory and readily discernible. In the case illustrated in Figure 9, it is apparent that effort should be devoted to improving hazard tracking practices, for example. Program initiation is seen at a marginal level, as well.

OPTIMIZING SCHEDULING

To serve advantageously at recognizing and ameliorating system safety program weaknesses, it is important that a measurement of program effectiveness be made at an optimum time in the lifecycle of an engineering program. This was recognized early in developing the model. If carried out too early, the system safety program may not have had an opportunity to become fully engaged. Applied too late, and the results of the model may be less effective in guiding mid-course improvements.

As an aid to resolving this matter, a poll was conducted among a small cross section of practicing system safety professionals. E-mail requests for expressions of judgment were sent to 15 currently active practitioners. Figure 10 reproduces the questionnaire that was used. Program phase descriptions from DoDI 5000.2 accompanied the questionnaire. (They appear here as Figure 11.) Nineteen responses were received and scored. (The population of responses exceeds the initial mailing; several poll participants shared ballots with colleagues who also responded.) Of the 19 respondents, 9 hold Professional Engineer (PE) certificates, Certified Safety Professional (CSP) or Certified Industrial Hygienist (CIH) certificates, or PhDs. Seven responded by telephone to participate in the poll. Four of these specifically requested anonymity. (All participants were assured of anonymity.)

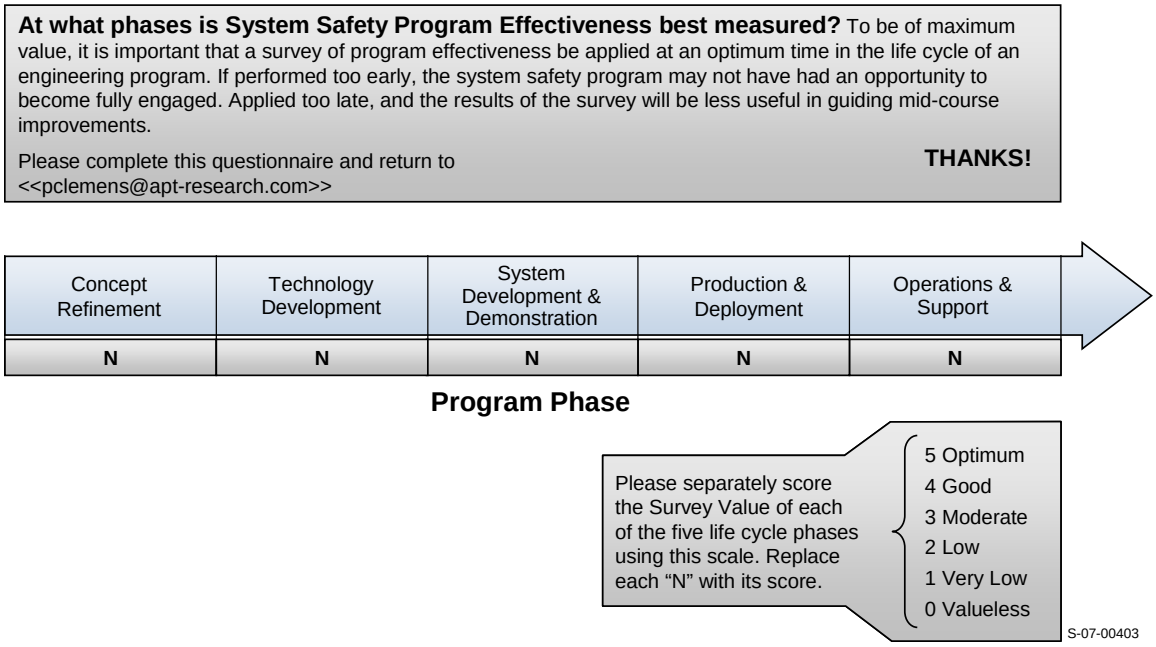


Figure 10. *Developing the Measurement Tool*

Concept Refinement: "The purpose of this phase is to refine the initial concept and develop a Technology Development Strategy."

Technology Development: "The purpose of this phase is to reduce technology risk and to determine the appropriate set of technologies to be integrated into a full system."

System Development & Demonstration: "The purpose of the SDD phase is to develop a system or an increment of capability; reduce integration and manufacturing risk (technology risk reduction occurs during Technology Development); ensure operational supportability with particular attention to reducing the logistics footprint; implement human systems integration (HSI); design for producibility; ensure affordability and the protection of critical program information (CPI) by implementing appropriate techniques such as anti-tamper; and demonstrate system integration, interoperability, safety, and utility."

Production & Deployment: "The purpose of the Production and Deployment phase is to achieve an operational capability that satisfies mission needs."

Operations & Support: "The objective of this activity is the execution of a support program that meets operational support performance requirements and sustains the system in the most cost-effective manner over its total life cycle."

*From Department of Defense Instruction Number 5000.2; "Operation of the Defense Acquisition System"

***Figure 11.** DODI 5000.2 Program Phase Descriptions**

Employment venues among the 19 respondents were distributed thusly:

- 7 DoD
- 5 NASA
- 7 Private Sector (all with current DoD and/or NASA contracts)

Five respondents contributed e-mailed "essays," some elaborating on their reasons for voting as they did, and others criticizing the poll. Their themes were devoted to such topics as these:

- "Why is the System Disposition (Disposal) phase not represented? This is [a serious] omission, especially [with] current concerns over environmental protection."
- "Concept Refinement" [earliest phase represented in the poll] is too late. "To be effective, System Safety should participate in concept selection and development, preceding concept refinement."
- "Measuring System Safety Program effectiveness should be done very early and then …repeated at intervals to guard against 'creeping degradation.' "

Poll data reduction was patterned after that used in reducing data from the interviews. The formula used appears in Figure 12, along with results of the poll. As the results show, application of the model to gauge system safety program efficacy is favored in the early program phases over the later phases by a very noticeable margin. Preference differences among the early three phases do not exceed 18%. A preference difference of only 3.5% separates the final two phases. However the average preference for the first three phases exceeds that for the final two by more than 35%. Among the five respondents who submitted unsolicited essays, two stressed the need to gauge program efficacy at multiple points during lifecycle.

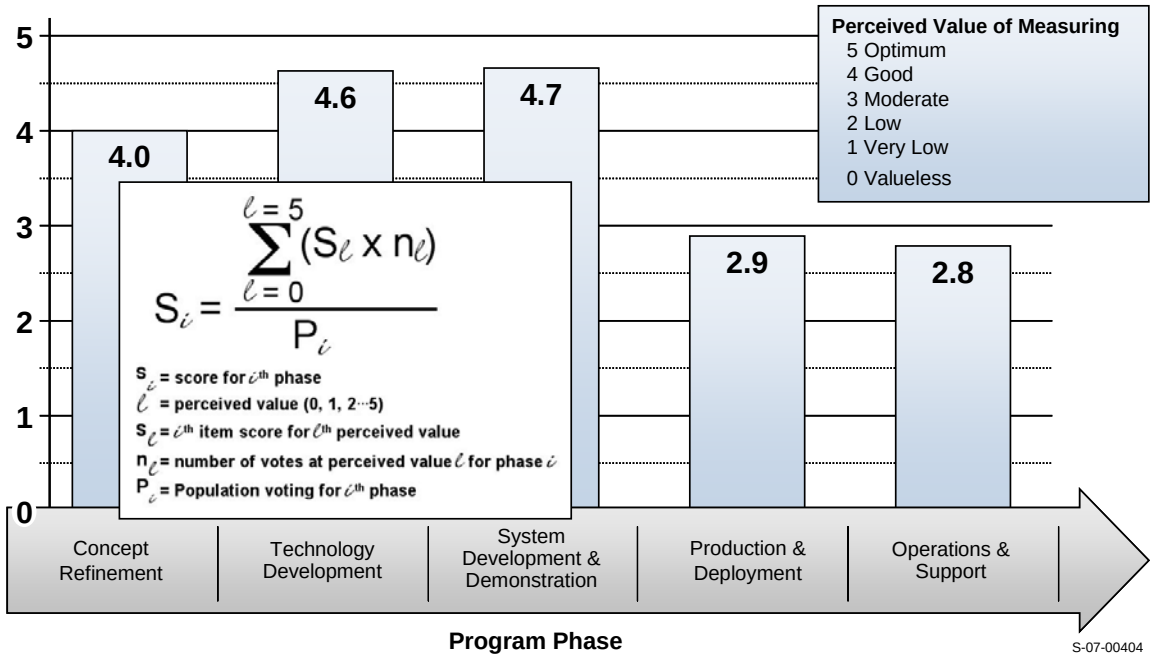


Figure 12. *Schedule Preferences*

A BETA TEST AND ITS RESULTS

Ms. Patricia Vittitow, Chief, U.S. Army Aviation and Missile Command (AMCOM) Safety Office, funded a beta test of the System Safety Performance Level Model as applied to a major DoD program. In this separately funded beta test, the model was applied by using interviews for data collection, as described above. Seventeen program practitioners from Government and contract personnel were interviewed. The interview sample was drawn from personnel having current and past direct involvement with the chosen program.

These lessons were learned from the beta test results:

- Cooperation by program management is an absolute necessity. Policies requiring evaluation may be necessary. (This was an observation by beta test Program Management. Full program management support was provided in the case of the beta test.)
- Fear that repercussion or reprisals can inhibit data accuracy. (This was an observation made by beta test Program Management. In the beta test, these concerns would have likely have actualized had there not been advance program management reassurances to "all hands" of absolute protection against disclosure of interviewee identifications with their contributions to results.)
- Understanding of the value-added aspect by Program Managers and safety professionals greatly improves acceptance.
- The necessary and essential data gathering interviews can be conducted by a safety engineer with no formal training or prior experience at interviewing for data gathering purposes of the kind involved here.
- The method has good potential as a program management guidance tool. (The beta test Program Manager declared it so in a formal panel briefing presented at the 2007 International Conference of the System Safety Society held in Baltimore, Maryland.)

- The method can pinpoint areas of needed system safety program improvements. (Same note as for item above.)
- The interview technique using a single interviewer is a cost effective data collection method. (Experience in the beta test verified that the use of a single interviewer is entirely adequate. Refs. 2 and 3 favor this approach as producing greater assurance of consistency of results.)
- Unless data scatter indicates otherwise, an interview sample size of 10% is adequate. (See Design Criterion 1, Table 1.)
- A randomly selected sample of program support practitioners serves well as a pool from which to gather the needed data.
- On average, formatted as at present, administering the interviews requires about 90 minutes each. This violates design criterion No. 1 of Table 1. Speeding the process sacrifices maintaining the collegial ambiance necessary to sustain the professionally motivated flow of information. (In the beta test, it became customary to conduct four closely scheduled interviews in one eight-hour day.)
- Data entry of information gathered in each interview requires no more than 10 minutes. (Data entries in the case of the beta test were made by a sophomore engineering co-op student. Data turn-around was described by Program Management as "brisk.")
- The model can be applied using multiple techniques (i.e, from self evaluation to full scale audit.)

The beta test verified the functioning of the method and demonstrated that "instant" results can be provided by its use. Evaluation results identified areas of needed improvement within the program. The model did not provide a comparison with a field-normed score nor provide assurance of proper application of analytical methods within the system safety program.

A FOLLOW-ON WORKSHOP

Representatives from the armed services and industry met on 05-06 June 2007 at the SEAC in Huntsville, Alabama to conduct the second of two workshops devoted to developing means for profiling the efficacy of system safety practice. A list of the 21 participants is at Appendix IV.

Workshop Goals

The goals of the workshop were to:

- Report on the development of the measurement "model" (i.e., progress made using results of the previous workshop, held in February, 2006).
- Unveil the model (draft).
- Report on the potential applications of the model.
- Describe the beta test and its results.
- Describe potential uses of data gathered using the model.
- Examine similar approaches.
- Discuss the way forward.
- Develop workshop consensus findings.

Workshop Summary

Presentations were made that discussed and described: development of the System Safety Performance Level Model; potential applications of the model; beta test results found by applying a particular application method; and system safety program efficacy measurement approaches already in use by others.

Achieving Accord/Consensus

Workshop members reviewed and commented on the original 37 inquiry items and associated measurement categories. The group determined that additional emphasis on Software Safety and Human Factors should be added to the questions. Minutes of this meeting a and a marked-up version of the questions was sent to participants for review and comment (Appendix V).

Workshop Findings

At the conclusion of the workshop, participants reached consensus on these points:

- The two workshops validated the model using a graybeard technique.
- A single index, carefully crafted for efficiency, can produce a useful measure of safety program performance – indicator of program efficacy.
- The model emphasizes technical aspects of the practice of the discipline rather than the administrative application or accuracy of results.
- Tailorable versions of the model for different types of systems may also provide benefit.
- The model is flexible to accommodate various system safety standards.
- The model is applicable to private sector, government industry, and government agencies.
- The model should be useful for identifying leading indicators for mishap prevention and product safety.
- Total ownership cost may be reduced if model-identified deficiencies are resolved.
- For software intensive systems additional emphasis on software may be necessary.
- The model is designed for and applicable to system safety programs independently of their size.
- The model could be implemented as a web-based system

The group determined that the model could be applied in various manners such as self assessments, independent assessments, and formal assessments. Ideally, safety programs should be assessed early and often for optimum results.

While the beta test effectively used the interview techniques there are other potential methods which could be used to obtain the necessary data. These include: program checklists, self-evaluations, questionnaires, in-depth surveys, and complete audits.

Metrics collected from the data could be used for benchmarking. Data gathered with sufficient fidelity could support multiple metrics. Metrics of interest include: element indices, composite index, index improvements, program trend evaluations, and safety experience, stability of safety organization, and safety culture.

SUMMARY

The system safety performance level model will serve as a useful tool to gauge the health of a safety program throughout the lifecycle of the program. Past experience has proved that a strong safety program results in significant savings to the program, reduced need for late application of corrective retrofits, and often more effective systems at lower overall cost. Just as early discovery and correction of safety program weaknesses favors cost savings, so does early opportunity for an effective safety program to influence design favor more effective system design outcomes at reduced overall cost.

REFERENCES

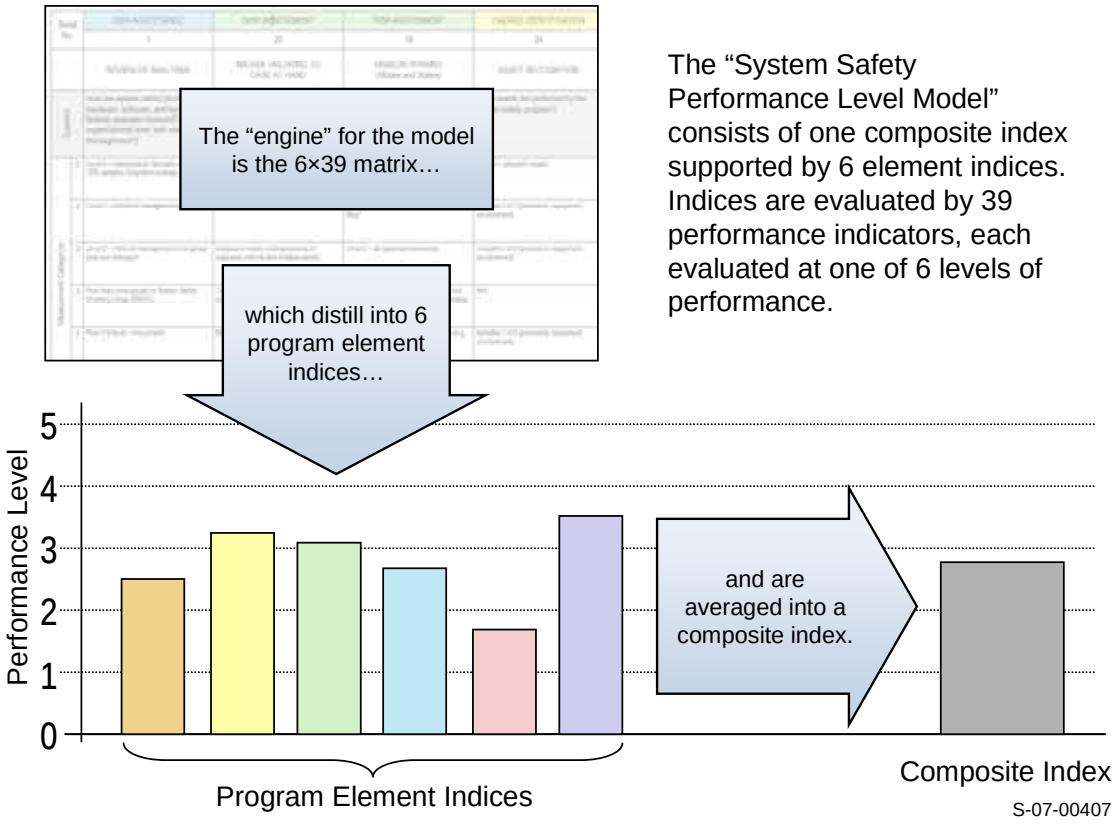
1. Sdorow, Lester M.; “Psychology”; McGraw-Hill; 1998; ISBN 0-697-25285-X
2. Ornstein, Michael D.; “Survey Research”; Current Sociology 46(4): iii-136; 1998 (as abstracted at <<http://en.wikipedia.org/wiki/Statistical_survey>>)
3. Scheuren, Fritz; “What is a Survey?”; taken from the American Statistical Association series of the same name and downloadable from <<http://www.whatisasurvey.info>>

Appendix I – The System Safety Performance Level Model

The System Safety Performance Level Model

The System Safety Performance Level Model depicted below consists of:

- Data Gathering Criteria
- Performance Scale
- Table of Inquiry Items
- Composite Index of Inquiry Items
- Results Database (Microsoft Excel spreadsheet)



Data Gathering Criteria

In gathering data for input into the model the criteria below are recommended.

Data Gathering Criteria

1	Interview ≥ 10% of program practitioners (increase sample size if data scatter indicates need)
2	Employ one-on-one interviews (one interviewer, one interviewee)
3	Protect Interviewee privacy (confidentiality must be assured)
4	Limit interviews to minimum number needed for evaluation—data scatter governs
5	Structure questions for discussion-style, rather than one-word responses
6	Explore both "traditional" and "cutting-edge" system safety practices

7	Evaluate responses according to multiple "performance determinants" developed and sanctioned by well-recognized practitioners
8	Interview duration ≈ one hour each
9	Provide adequate resolution of results (i.e., ≥ 5 discrete measurement levels)
10	Provide results interpretable as guidelines for System Safety Program improvement (if needed)

Performance Scale

The performance scale shown below used to score responses to questions to gauge the "goodness" of the relevant element is system safety practice. Scores range from the extreme of absent practice to an opposite superlative.

Performance Scale

Score	Descriptor
5	Excellent: Superlative in both concept and performance
4	Good: Satisfying all needs, in letter and in spirit
3	Fair: Satisfying most needs, but with one or a few notable flaws
2	Marginal: Failing to satisfy one or more key needs
1	Minimal: Present, but having little or no value to the Program
0	Null or Absent: Virtually non-existent/not practiced

Inquiry Items

The questions making up the central body of the model appear below. Also shown there is the system safety program element with which each question is associated.

Inquiry Items

#	Question	Program Element*
1	How are system safety (including hardware, software, and human factors) analyses reviewed? (At what organizational level; with what thoroughness?)	5
2	How are waivers, exceptions, and other non-compliances managed by the system safety program?	5
3	How does system safety (including hardware, software, and human factors) manning allocation compare to actual needs?	1
4	Are the six program elements and functions recognized? (i.e., program initiation, hazard identification, risk reduction, risk acceptance, hazard tracking)	1
5	What authority does the system safety organization have?	1
6	Which model best describes system safety organization structure and reporting level?	1
7	Is management practice consistent with current standards?	1
8	What level of regard prevails for the system safety program description documents (including hardware, software, and human factors)?	1

#	Question	Program Element*
9	Does the system safety program specifically address Environmental, Safety, and Occupational Health (ESOH) hazards?	1
10	How is system safety practice linked to the "illities" – e.g., Reliability, Availability, Maintainability?	6
11	How often, long-term average, has the program made system safety organizational/personnel changes?	1
12	How is acceptance of residual risk documented/reported?	5
13	Are well-trained system safety personnel involved in the specification and design process? Do the design specifications reflect the impact of safety inputs?	1
14	How are hazard analyses results tracked, recorded, documented, updated, and closed?	3
15	How is risk assessed for a single hazard that threatens multiple protected assets? Use clarifying example.	3
16	How are risk tolerance limits selected for programs?	5
17	What system safety risk summation practices are employed?	5
18	How is mission phasing (modes and states) recognized in system safety (including hardware, software, and human factors) analysis?	3
19	How is exposure interval selection made?	3
20	How is risk assessment matrix tailoring done?	3
21	In assessing risk, how is uncertainty addressed or characterized?	4
22	How is the hierarchy of mitigation precedence treated?	4
23	Are risk assessments quantified?	4
24	What assets are protected by the system safety program?	2
25	How are hazard tracking and safety requirement traceability implemented for system safety (including hardware, software, and human factors)?	6
26	How are hazards identified?	2
27	What hazard inventory-type analytical techniques are used?	2
28	What logic tree analytical techniques are used?	2
29	Are safety hazard data maintained up-to-date?	6
30	How are hazards understood by those working in the program?	2
31	Is analytical tool selection tailored to the needs of individual system peculiarities and needs?	3
32	Are newly discovered hazards reported and mitigation plans formulated promptly?	4
33	What level of training in system safety (including hardware, software, and human factors) have you achieved?	1
34	How many years of direct, full-time equivalent experience have you had in system safety (including hardware, software, or human factors) practice?	1
35	What applicable credentials do you hold?	1
36	How would you characterize the system-specific knowledge of system safety staff members?	1
37	How would you characterize organizational safety culture?	1

#	Question	Program Element*
38	For software-intensive systems, how are the software personnel involved as contributors in identifying the hazards and mitigators?	2
39	How is the analysis of Government-off-the-Shelf (GOTS)/Non-Developmental-Items (NDI)/Commercial-off-the-Shelf (COTS)/software reuse addressed in the system safety program (including hardware, software, and human factors)?	3

* 1 Program Initiation	2 Hazard Identification	3 Risk Assessment	4 Risk Reduction	5 Risk Acceptance	6 Hazard Tracking

Data Collection Sheets

Data collection sheets can be implemented in various ways. If the interview option is chosen to implement the model, the sheets can be used in the interview process. The sheets list the system safety element from which the question comes and the question and measurement categories for each. Inputs from the sheets may be entered into the results database to produce graphical displays of results.

Serial No.		RISK ACCEPTANCE	RISK ASSESSMENT	RISK ASSESSMENT	HAZARD IDENTIFICATION
		1	20	18	24
		REVIEW OF ANALYSES	MATRIX TAILORING TO CASE AT HAND	MISSION PHASING (Modes and States)	ASSET RECOGNITION
Question		How are system safety (including hardware, software, and human factors) analyses reviewed? (At what organizational level; with what thoroughness?)	How is risk assessment matrix tailoring done?	How is mission phasing (modes and states) recognized in system safety (including hardware, software, and human factors) analysis?	What assets are protected by the system safety program?
Measurement Categories	5	Level 4 + independent, 3rd party review of >5% samples, long-term average	Full matrix (indices/spans/resolution) quantitatively tailored, asset-by-asset.	Level 4 + maintenance/calibration, repair, other support phases	Level 4 + program impact
	4	Level 2 + 2nd level management or above	Quantitative partial matrix scaling/tailoring.	Level 3 + contingencies, e.g., "Emergency Stop"	Includes 3 of 3 (personnel, equipment, environment).
	3	Level 2 + 1st level management (one group plus one manager)	Subjective matrix scaling/tailoring of exposure interval and multiple assets.	Level 2 + all significant transients	Includes 2 of 3 (personnel, equipment, environment).
	2	Peer team (one group) or System Safety Working Group (SSWG)	Tailored severity scale for only one asset (e.g., equipment or personnel).	Tailored to obvious operating phases, but omitting transients (i.e., activities separating major functions).	N/A
	1	Peer (1st level - one person)	Need recognized, but none performed.	Modest, pro-forma, major phases only (e.g., startup/run/stop).	Includes 1 of 3 (personnel, equipment, environment).
	0	None performed.	Need not recognized.	Not recognized.	No distinctions made.

Serial No.		RISK ASSESSMENT	HAZARD IDENTIFICATION	RISK REDUCTION	RISK REDUCTION
		15	30	22	21
		ASSET RISK ASSESSMENT	UNDERSTANDING/DESCRIBING HAZARDS	MITIGATION PRECEDENCE OBSERVATION	HANDLING UNCERTAINTY
Question		How is risk assessed for a single hazard that threatens multiple protected assets? Use clarifying example.	How are hazards understood by those working in the program?	How is the hierarchy of mitigation precedence treated?	In assessing risk, how is uncertainty addressed or characterized?
Measurement Categories	5	Risk is individually assessed for each of multiple assets.	Thorough understanding: a hazard constitutes a threat of harm to one or more assets and is expressed as a source, a mechanism, and an outcome.	Level 4 + proper use of design change is generously evident.	Uncertainty is evaluated with rigor, on a case-by-case basis, according to a prescribed, documented plan.
	4	Assessments for multiple assets are aggregated using a defined rule.	N/A	Level 3 + requirement for use is documented/enforced.	Uncertainty is assessed either subjectively or quantitatively but without a standardized, documented plan.
	3	Assessments for multiple assets are aggregated consistently by a rule developed by the individual practitioner.	Hazards are moderately well understood but are described inconsistently and don't specify specific threatened assets.	Hierarchy is properly used, and use is monitored/reviewed.	"Standard" uncertainties are assigned to subjective severity and probability evaluations then projected to a value for risk.
	2	Assessments for multiple assets are "lumped" to a single risk declaration without disciplined aggregation.	N/A	Hierarchy is properly used, but use is ill-enforced.	Risk uncertainties are subjectively judged from a standardized scale as are levels of hazard probability and severity.
	1	Risk is assessed/reported only for the single asset which is judged to have greatest risk.	Hazards are often described, simply as a source, or a mechanism, or an outcome, alone.	Hierarchy is recognized but use is not monitored or enforced; mitigation measures are often mis-ranked.	No consideration is given to evaluating uncertainty, or uncertainty is very poorly conceptualized.
	0	Assessment method not known or not understood by interviewee.	Widespread misunderstanding as to what constitutes a hazard.	Effectiveness hierarchy not recognized, not used.	Uncertainty as a concept is inadequately understood to be applied.

Serial No.		HAZARD TRACKING	PROGRAM INITIATION	HAZARD TRACKING	HAZARD IDENTIFICATION
		25	13	10	26
		HAZARD TRACKING	SYSTEM SAFETY INFLUENCE ON DESIGN (Hardware, Software, Human Factors)	CROSS-COUPLED "ILLITIES"	HAZARD IDENTIFICATION
Question		How are hazard tracking and safety requirement traceability implemented for system safety (including hardware, software, and human factors)?	Are well-trained system safety personnel involved in the specification and design process? Do the design specifications reflect the impact of safety inputs?	How is system safety practice linked to the "illities" – e.g., reliability, availability, maintainability?	How are hazards identified?
Measurement Categories	5	Level 4 + auditable evidence that the hazard has been mitigated to an acceptable level of risk and there is an audit trail of safety requirements.	Designers are trained in system safety; immediate application of system safety principles is evident. Safety lessons learned considered.	Full-bore, readily-auditable linkage to reliability, availability, and maintainability.	Formally prescribed balance of brainstorm, checklists, walkthroughs, and hazardous operations (HAZOP), Failure Modes and Effects Analysis (FMEA), or Failure Hazard Analysis (FHA).
	4	Level 3 + coupled with configuration, management, or quality program.	Major influence in specifications and through practice of concurrent engineering or equivalent.	Formal, mandatory cross-feed with reliability and another "illity."	Level 3 + use of HAZOP, FMEA, or FHA, or equivalent; or use of prior experience with like systems.
	3	Procedure-driven and documented in well maintained records, uniform format, and with a well established process. Safety requirements derived from hazards.	Influence through safety participation in determining specifications and in design reviews.	Formal, mandatory cross-feed with reliability or availability, or maintainability.	Level 2 + supported by checklist(s) and/or energy source inventory and/or operational walkthroughs.
	2	Practiced according to loosely interpreted standards and procedures.	Influence through safety participation in infrequent design reviews.	Modest, moderately formal cross-feed with reliability or availability, or maintainability.	Organized, formally led brainstorming.
	1	Informally practiced.	Modest influence through inconsistent and infrequent design reviews.	Infrequent, informal cross-feed with reliability or availability, or maintainability.	Informally guided brainstorming; "what if."
	0	Not practiced.	Little or no evidence of influence on design.	Not linking practiced.	No formally required or documented techniques.

Serial No.		RISK ACCEPTANCE	RISK ACCEPTANCE	RISK ACCEPTANCE	HAZARD IDENTIFICATION
		2	16	17	27
		MANAGING WAIVERS	RISK TOLERANCE LIMIT SELECTION	RISK SUMMATION	HAZARD INVENTORY TOOLS
Question		How are waivers, exceptions, and other non-compliances managed by the system safety program?	How are risk tolerance limits selected for programs?	What system safety risk summation practices are employed?	What hazard inventory-type analytical techniques are used?
Measurement Categories	5	Waivers are time-limited and are tracked to soundly justified resolution and closeout without renewal.	Hazard probability and severity levels, exposure interval, and risk acceptance contour tailored by management to satisfy program needs; requirement is documented and enforced.	Risk summation is required and enforced with rigorous quantitative calculations and software support; total system risk versus partial risk is well recognized.	One or more top-down methods and one or more bottom-up methods, formally documented, with software databasing.
	4	Waivers are properly requested /approved, and tracked, but occasionally are insufficiently time-limited, or closeout deadlines are extended with little question/no stated justification.	Tailored by adjusting hazard severity and hazard probability level definitions, adjusting risk tolerance contours within matrix, and adjusting exposure interval.	Summing risks is required; quantitative calculations are well understood, and the concept of total risk versus partial risk is well recognized.	Formally required FMEA or FHA, or equal; requirement is documented.
	3	Waivers are properly requested/approved but many persist indefinitely, or are arbitrarily closed out with questionable justification (e.g., "it hasn't happened yet, so it's not a hazard").	Tailored by adjusting hazard severity and/or probability scale definitions or exposure interval.	Summing risks is a required practice; subjective application is widely recognized, and quantitative application is moderately understood.	Formally required Preliminary Hazard Analysis (PHA) or HAZOP, with tailored matrix use; requirement is documented.
	2	Waivers are requested/approved perfunctorily, without probing the rationale for granting them and are arbitrarily closed or persist indefinitely.	Applied directly from a carefully selected standard (e.g., MIL-STD-882), with no tailoring, with fixed, documented exposure interval.	Risk summation is moderately understood; subjective application is often used, but quantitative calculation is poorly understood.	Formally required PHA, without tailored matrix use; requirement is documented.
	1	Waiver practice requirements are very poorly understood by line personnel and/or are poorly documented/implemented.	Pro-forma, directly from a standard (e.g., MIL-STD-882) without change and without regard for exposure interval.	Risk summation is modestly understood; the concept is loosely interpreted and practiced subjectively.	Formally required PHL, requirement is documented.
	0	Program has no provisions for waivers - none are used.	Risk tolerance concepts are not recognized; standards-based code worthiness is the sole risk tolerance determinant.	Risk summation is insufficiently understood to be used by the program; partial risks are treated individually.	Informal brainstorm list-making.

Serial No.		HAZARD IDENTIFICATION	RISK ASSESSMENT	PROGRAM INITIATION	PROGRAM INITIATION
		28	19	6	5
		LOGIC TREE TOOLS	EXPOSURE INTERVAL SELECTION	STRUCTURE/REPORTING	AUTHORITY
Question		What logic tree analytical techniques are used?	How is exposure interval selection made?	Which model best describes system safety organization structure and reporting level?	What authority does the system safety organization have?
Measurement Categories	5	Probabilistic risk assessment, fully quantified with sound understanding of uncertainty.	Exposure Interval is explicitly stated, with due regard for the overall time interval and adjusted for less-than-full-time functions.	Staff organization reporting to general manager and/or program manager and has access to Integrated Product Teams (IPTs) (or equivalent).	"Must Change" and "Stop Work" authority documented and enforced.
	4	Level 3 + Cause-Consequence Analysis (CCA) or equal (quantified) with a very sound understanding of calculating risk from probability and severity assessments.	Exposure Interval is explicitly stated, with due regard for the overall time interval, but not adjusted for less-than-full-time functions.	Integrated with engineering and reporting to general manager.	"Stop Work" or "Must Change" authority (in redesign), with evidence of application.
	3	FTA and/or Event Tree Analysis (ETA) with a very sound understanding of probability assignments and calculations.	Exposure Interval is explicitly stated, but mis-calculated.	Staff organization reporting to an engineering organization.	Moderate design change authority, often overruled.
	2	Level 1 + ETA (unquantified) (or another logic tree method) with a sound understanding of modeling events with binary outcomes.	Exposure Interval is known to play a role in determining Hazard Probability but is unstated or is assigned from a standard, not adjusted to true Program needs.	Staff organization reporting to an operations organization with regular informal communication with the engineering organization.	Formal advisory.
	1	FTA (unquantified) with a sound understanding of "AND" and "OR" logic gates.	Exposure Interval is vaguely/indefinitely stated (e.g., "life cycle").	Staff organization reporting to an operations organization.	Observe and comment.
	0	No formal logic tree analytical techniques are used.	Need for expressing an explicitly stated Exposure Interval is not recognized.	Staff organization reporting to Human Resources (HR) or other similar non-technical organization.	None.

Serial No.		PROGRAM INITIATION	PROGRAM INITIATION	PROGRAM INITIATION	PROGRAM INITIATION
		33	34	35	3
		SPECIALIZED TRAINING	EXPERIENCE	FORMAL CREDENTIALS	DEDICATED LoE vs NEEDS
Question		What level of training in system safety (including hardware, software, and human factors) have you achieved?	How many years of direct, full-time equivalent experience have you had in system safety (including hardware, software, or human factors) practice?	What applicable credentials do you hold?	How does system safety (including hardware, software, and human factors) manning allocation compare to actual needs?
Measurement Categories	5	Level 3 + over ten years domain knowledge.	>25 years direct experience in system safety or equivalent domain knowledge.	Professional Engineer (PE) certification or advanced degree in engineering, physics, math, computer science, or related discipline + System Safety Society member (professional grade).	Adequately staffed; very rare overtime need; no schedule overruns (>5% of total engineering staff).
	4	Level 3 + five years applicable domain knowledge.	15-25 years direct experience in system safety or equivalent domain knowledge.	Level 3 + Certified Safety Professional (CSP) or Associate Safety Professional (ASP).	Adequately staffed; infrequent overtime need; very rare schedule overruns (≈4-5% of total engineering staff).
	3	Level 1 + formal classroom training (≥30 classroom hrs) specifically in system safety engineering, with CEUs or college credit.	7-15 years direct experience in system safety or equivalent domain knowledge.	Level 2 + System Safety Society member (professional grade).	Marginally understaffed; some overtime need; few schedule overruns (≈3% of total engineering staff).
	2	Level 1 + formal classroom training (≥20 classroom hrs) specifically in system safety engineering.	3-7 years direct experience in system safety or equivalent domain knowledge.	Bachelor of Science (BS) in engineering, physics, math, computer science, or related discipline.	Noticeably understaffed; some overtime need; some schedule overruns (≈2-3% of total engineering staff).
	1	One year or more of on-the-job training (i.e., *not* one year of identical methodology/assignment).	1-3 years direct experience in system safety or equivalent domain knowledge.	High School diploma + specialized training in system safety (w/certificate).	Unquestionably understaffed; frequent overtime need; frequent schedule overruns (≤1% of total engineering staff).
	0	No formal system safety training and less than 1 year of on-the-job training.	< 1 year direct experience in system safety or equivalent domain knowledge.	High school diploma or GED with no specialized training in system safety.	Unable to perform properly; e.g., need not recognized, overtime disallowed, not incl. in schedule or budget (<1% of total engineering staff).

Serial No.		PROGRAM INITIATION	PROGRAM INITIATION	PROGRAM INITIATION	HAZARD TRACKING
		36	37	4	29
		SYSTEM-SPECIFIC KNOWLEDGE	SAFETY CULTURE	SIX PROGRAM ELEMENTS	SAFETY DATA CURRENCY
Question		How would you characterize the system-specific knowledge of system safety staff members?	How would you characterize organizational safety culture?	Are the six program elements and functions recognized? (i.e., program initiation, hazard identification, risk assessment, risk reduction, risk acceptance, hazard tracking)	Are safety hazard data maintained up-to-date?
Measurement Categories	5	Full program knowledge; safety staff members have a well developed understanding of the entire system and its requirements, general and specific.	Demonstrated, full corporate involvement; system safety a highly esteemed part of any project.	All six are recognized, documented, and practiced.	Updating required and practiced in response to: mishaps/near misses, design changes, or progress in modeling, analysis, or simulation.
	4	A well developed familiarity with the entire system and its requirements; capable of providing technical support in many system-specific areas.	Assigned management is knowledgeable and effective; system safety contributions are well-appreciated.	The six discrete elements are practiced but without clear System Safety Program Plan (SSPP) documentation.	New experiences with same or similar systems prompt updating.
	3	Moderately developed familiarity with the entire system and its requirements; good understanding of selected system subparts.	System safety considered a part of systems engineering or on a similar engineering level; management is moderately well-informed of the system safety role.	Six elements recognized in program plan but poorly practiced (practice under-enforced or need under-appreciated).	Major developments or findings of potential loss events, including near misses and incidents, prompt updating.
	2	Introductory familiarity with part of the system; moderate familiarity with some system specifics.	System safety is considered below engineering in importance or authority and is underappreciated, but performs notable function, though hampered.	Some recognition apparent in program plan documentation, but six elements not fully practiced.	Major and minor loss events prompt updating; updating is not always timely.
	1	Introductory familiarity with part of the system; little or no familiarity with overall system specifics.	System safety role is poorly defined /understood; engineers have negative bias toward system safety personnel and function.	Elements not documented in program plan; only a few elements recognized, and poorly practiced.	Major loss events prompt updating; updating is not always timely.
	0	System knowledge poorly developed; personnel poorly informed or lacking in technical proficiency.	Culture undeveloped or requirement not understood by program personnel; "system safety" function unrecognized or viewed negatively by many.	The six elements are not recognized explicitly or in practice; no documentation specifies them as discrete program elements.	Data are rarely updated or not updated at all; updating requirements do not exist.

Serial No.		RISK ACCEPTANCE	RISK REDUCTION	RISK ASSESSMENT	RISK ASSESSMENT
		12	32	14	31
		RESIDUAL RISK ACCEPTANCE DOCUMENTATION	PROMPT HAZARD REPORTING/ MITIGATION	ANALYSIS DOCUMENTATION	TAILORED TOOL SELECTION
Question		How is acceptance of residual risk documented/reported?	Are newly discovered hazards reported and mitigation plans formulated promptly?	How are hazard analysis results tracked, recorded, documented, updated, and closed?	Is analytical tool selection tailored to the needs of individual system peculiarities and needs?
Measurement Categories	5	Signed risk assessment document, electronically archived in searchable database.	Discovery/hazard reporting and risk assessment by collaborating safety/engineering team; prompt mitigation development/implementation.	Recorded in uniform format software with search function and software-compiled summaries accessible by all system safety personnel.	Optimized selection of tools according to need from wide range of options.
	4	Signed risk assessment document, with archived paper copy.	Hazard/risk assessment reporting by system safety personnel with immediate reporting to engineering and/or Program Management (PM) level; mitigation follows shortly.	Recorded with uniform format for entire program in software with search/sort capability.	Tools generally selected to fit the task, but sometimes chosen according to analyst's capability rather than system need.
	3	Verbal (spoken), transcribed/documented in hazard tracking software.	Infrequent reporting at periodic meetings dedicated to that purpose; mitigation follows, usually with no particular urgency.	Recorded using software with database search/sort capability.	Moderate tailoring of selection from standard list of tools for all systems/subsystems.
	2	Verbal (spoken) transcribed informally on handwritten sheet.	Reporting done only during widely spaced formal design reviews; mitigation follows.	Recorded using software without database search/sort capability.	No tailoring; standard, modest selection of tools for all systems/subsystems.
	1	Verbal (spoken) with no documentation of residual risk having been accepted.	Reporting significantly lags discovery, delaying mitigation.	Handwritten data sheets.	One tool used without modification for all analyses.
	0	No recognition or flawed recognition of residual risk as a concept.	No formal reporting method exists; mitigation of a newly discovered hazard may not occur until a near miss or loss event is experienced.	Analysis documentation is not developed; no requirement exists for documentation or its method.	Need for tailored selection not recognized.

Serial No.		PROGRAM INITIATION	PROGRAM INITIATION	RISK REDUCTION	PROGRAM INITIATION
		7	9	23	11
		CONSISTENCY WITH CURRENT SYSTEM SAFETY STANDARDS	TREATMENT OF ESOH HAZARDS AND RISKS	RISK ASSESSMENT QUANTIFICATION	ORGANIZATIONAL STABILITY
Question		Is management practice consistent with current standards?	Does the system safety program specifically address Environmental, Safety, and Occupation Health (ESOH) hazards?	Are risk assessments quantified?	How often, long-term average, has the program made system safety organizational/personnel changes?
Measurement Categories	5	Management practice is fully consistent with applicable current standards; measures are in place to remain consistent and current.	Hazards of occupational injury/illness and environmental hazards are recognized, and their risks assessed.	All risk assessments are quantified.	Very rarely — less than once in four years.
	4	Mostly consistent, with inconsistencies only in minor areas or during changes in standard version.	Hazards of occupational injury/illness and environmental hazards are usually recognized but their risks are often mis-assessed.	N/A	Rarely — less than once in three years.
	3	Regularly inconsistent in minor matters and occasionally inconsistent in key matters; widespread understanding of standards but lack of full enforcement.	Hazards of occupational injury/illness and environmental hazards are usually recognized but only for hazards also threatening personnel or equipment.	Some risk assessments are numerically expressed.	Occasionally — less than once in two years.
	2	Moderate inconsistency in key matters; lack of universal understanding/enforcement of standards.	ESOH hazards are moderately well recognized.	N/A	Often — about once in two years.
	1	Markedly inconsistent, with evidence of deliberate neglect or ignorance of applicable standards.	ESOH hazards are poorly recognized.	An underdeveloped effort has been made to numerically express/quantify risk.	Frequently — about once a year.
	0	Entirely inconsistent; little or no understanding of applicable standards.	ESOH hazards are not recognized.	No risk assessments are numerically expressed.	Very often — more than twice a year.

Serial No.		PROGRAM INITIATION	HAZARD IDENTIFICATION	RISK ASSESSMENT
		8	38	39
		PROGRAM PLAN STATUS	SOFTWARE SAFETY INVOLVEMENT	ASSESSMENT OF LEGACY ITEMS
Question		What level of regard prevails for system safety program description documents (including hardware, software, and human factors)?	For software-intensive systems, how are the software personnel involved as contributors in identifying the hazards and the mitigators?	How is the analysis of Government-off-the-Shelf (GOTS)/Non-Development-Items (NDI)/Commercial-off-the-Shelf (COTS)/software reuse addressed in the system safety program (including hardware, software, and human factors)?
Measurement Categories	5	Program plan is a "Living Document," referred to frequently, updated as necessary, and used as a baseline guide to program operation.	Fully involved in System Safety Working Group (SSWG) and integrated into the various IPTs.	Fully defined with appropriate variety of techniques for certification of proposed use, monitoring and selection.
	4	Program plan is used as a guide in most program operations but is in need of being updated.	Level 3+ involvement in developing mitigation features.	Limited safety approval process.
	3	Program plan exists, served to initiate program activities, but is no longer referred to or used to guide program operations.	Involved in SSWG.	Monitoring safety of use.
	2	Program plan is a "Dead Document," rarely referred to, and with many disregarded or outdated provisions.	Involved in IPTs (or equivalent) only.	Defined using appropriate techniques.
	1	If a program plan exists, practitioners are unacquainted with its provisions.	Involved in limited IPTs (or equivalent) only.	Defined methods for item selection.
	0	No program plan exists.	No involvement.	Not addressed.

Appendix II – Good System Safety Practices Workshop Attendance

21-22 February 2006

Name	Organization	Phone	21 Feb	22 Feb
Arun Murthi	Aero & Space USA	714-267-6888	√	√
Barry Hendrix	LMCO	770-494-6961	√	√
Bill Edmonds	U.S. Army CRC	334-255-1122	√	√
Bob Baker	APT Research	256-327-3371	√	√
Bob McAllister	WPAFB	937-257-0470	√	√
Craig Schilder	NexPort	703-399-6520	√	√
Dave West	SAIC	256-971-6494	√	√
Dev Raheja	DCI	301-483-4525		√
Don Swallom	U.S. Army AMCOM	256-842-8641	√	√
Herb Shivers	NASA	256-544-8903	√	√
J.R. Rao	OID/BIW	207-442-1494		√
James Inge	MOD-UK	+44 117 913 5739	√	√
James Taylor	Westar	334-255-2744	√	√
James Wiggins	Raytheon	256-542-4289	√	√
Jim Gibbons	Boeing	610-591-8813	√	√
John Frost	Safety Engineering Services	256-650-0335	√	√
John Leipper	MCPD	760-731-3472	√	√
John McDermid	University of York	+44 190 443 2726	√	√
Lucio Tolentino	Aerospace Corporation	310-336-6723	√	√
Mark Geiger	U.S. Navy	703-602-5020	√	√
Mary Ellen Caro	NOSSA	301-744-6039	√	√
Mike Wesoloski	LMSSC	408-742-0172	√	√
Paige Ripani	Booz Allen Hamilton	703-412-7702	√	√
Pam Kniess	PEO Aviation	256-828-5467	√	√
Pat Clemens	APT Research	256-327-3707	√	√
Philip Smiley	U.S. Navy	703-693-4017	√	√
Rod Simmons	Illinois State University	309-438-7133	√	√
Ron Howlett	MOD-UK	+44 117 913 5381	√	√
Saralyn Dwyer	APT Research	256-327-3377	√	√
Sherman Forbes	SAF/AQRE	703-588-7839	√	√
Sid Smith	APT Research	256-327-3397	√	√
Tom Pfitzer	APT Research	256-327-3388	√	√
Tomoso Sgobba	ESB	+3 1715 654 966	√	√

Appendix III – Guidance for the Interviewer Using the System Safety Performance Level Model

POINTS FOR DISCUSSION AT THE OUTSET OF THE INTERVIEW

Before beginning the interview, make the interviewee comfortable and provide background information that will help the interviewee understand the rationale underlying the interview. Here are some brief discussion points that will help to accomplish this. If preferred, this information may be provided to the interviewee as a handout a day or two before the interview. However, its use as a conversation opener at the location and time of the interview is helpful in establishing a sense of professional rapport with the interviewee. Its value in this role should not be overlooked.

Measurement Method: The purpose of the System Safety Performance Level Model is to gauge the efficacy of system safety program performance. To accomplish this, interviews are conducted with program practitioners. Answers to standardized interview questions provide data that are then used to profile program performance. Questions used in the interview are based on initial work done in Huntsville, Alabama in February 2006, during a two-day workshop. Workshop participants were a group of more than forty System Safety specialists from around the US and abroad. Participants came from private industry, DoD, NASA, the European Space Agency, several universities, and the FAA. The design of the interview followed from the workshop activity. A beta test of the technique followed. Beta test results were reviewed in a follow-on two-day workshop, also held in Huntsville, in May 2007. Participants there guided several improvements. These have now been incorporated into the interview material.

The Purpose in gathering information in the interview is not to find fault, but to gauge program performance attributes — to identify particular areas of program performance strength and weakness. The overall goal is to evaluate program effectiveness as a basis for guiding improvement in performance areas where improvement may be desirable.

Anonymity: Interviewees' responses to questions will be wholly protected against disclosure as to source. Interviewees should feel free to provide completely candid responses.

Duration: Experience has shown that the usual interview consumes about 90 minutes. A break can be taken at any time either participant, the interviewee or the interviewer, might wish.

THE INTERVIEW

PREPARATORY STEPS

In advance of beginning the interviews, spend some time familiarizing yourself with the contents of the questions you'll be asking the interviewee and the scoring scheme used in collecting results. Read through the interview questions to ensure that you'll be comfortable in asking them during the interviews. Do not let the contents of the interview take you by surprise during the course of the interviews.

Closely related topics are often addressed by questions that are separated by unrelated ones in the interview series. The separation gives an impression of disorderliness. This arrangement is intentional however. If items bearing on the same or on a closely related topic are closely spaced, an emotionally driven quick response to one of them may unduly influence responses to the following one(s). Intervening questions dealing with unrelated topics provide a "cooling off period to promote responses that are more likely to be reflective than emotionally driven.

Scheduling: Set a specific time for each interview. Experience has shown that the average interview can be conducted in about 90 minutes. Scheduling interviews at two-hour intervals works well. Be present at the appointed hour. Do not make the interviewee wait!

Conduct the interview in a "neutral" location rather than in your or the interviewee's workplace setting. Select a quiet location that affords privacy, one with comfortable seating and a table. Do not sit immediately opposite the interviewee. Doing so conveys a confrontational image. If the interviewee selects a position at a side of the table, seat yourself at the nearest end, and vice versa.

Things to bring to the interview should include:

a. One copy of the complete set of interview questions. You'll use it as the question source, and you'll use its example responses (the "Measurement Category" entries) as guides to be used in scoring. If scores are marked directly on the set of interview questions, using one set for each interviewee, recordkeeping is then simplified. CAUTION: In no case should the interviewee's name appear on the scored set of questions.

b. A clipboard is a convenience for manipulating the interview question sheets.

c. Two copies of the "Performance Scale." One of these is to be given to the interviewee. You and the interviewee will work as a team in arriving at question-by-question scores.

d. A writing pad on which to make notes on observations of matters not covered in the interview question sheets.

CONDUCTING THE INTERVIEW

Carry out the interview one-on-one, with no one else present. Presence of others can

intimidate the interviewee or create an atmosphere conducive to confabulation and "acting out."

Pose the interview questions in whatever manner is comfortable for both you and the interviewee. Paraphrasing is permitted and is preferable to a "dry" verbatim recitation so long as the sense of the question remains intact. It is the basic, program-related, performance factor embodied in each query that is important, not delicate nuances.

Avoid hurrying, or any appearance of it, but maintain a comfortable pace without giving way to protracted recitations of "war stories" by yourself or the interviewee.

SCORING RESPONSES

Provide a copy of the Performance Scale to the interviewee. Explain that it will serve as the guide for scoring system safety program attributes, as seen by the interviewee.

Score interviewee responses without the expectation of capturing "exact matches" with the example responses that are provided. The example responses are meant to apply to general rather than to specific cases. They are shown only to provide a sense of scaling. Resolve cases of doubt by using guidance from the Performance Scale.

Be objective and non-judgmental. Do not allow the interviewee to sense any personal views you may have as to program merit. Remain stolidly objective and professional in your personal demeanor, but not unfriendly. Keep in mind those things that reveal viewpoint and attitude, however subtly, e.g.: inflection of voice, selection of vocabulary, "body english," facial expression.

Facilitate but do not guide responses. Your goal is to elicit a factual response to each question, a response free of any influence of emotional bias that might be injected by either you or the interviewee. (Providing the interviewee the copy of the Performance Scale aids in this. Let the interviewee recognize that the two of you are a team of system safety professionals, seeking the truth.)

Allow time for the interviewee to reflect and respond. (See Item 8, above.) Discount "knee-jerk" responses given too hastily, whether they are positive or negative. Quickly given responses are often emotionally driven and are likely to exaggerate or to stray from the truth. Discussion-style responses will be more thoughtfully considered. To elicit accuracy, encourage discussion.

Gently deflect too-quick responses without interrupting the flow of the interview and without signaling either agreement or disagreement with a position expressed by the interviewee. Often this can be accomplished by using flatly expressed, noncommittal comments, such as:

a. "Tell me a little about what's behind that, please." [This will provide "pause time" for reflection.] — or,

b. "Is that something you'd also have said about other programs you've worked on, or do you think it's pretty rare?" [This can serve to prompt contemplating a sense of "scale."] — or,

c. "Let's think about just where that'll put us on the Performance Scale for this question." [Again, this provides an opportunity for reflection and scaling.]

"Bargain" to score tough questions, but bargain with the Performance Scale *not* with the interviewee! As a two-person team of professionals, yourself and the interviewee, negotiate with the applicability, of the levels of the scale to the item under consideration. This can be a particularly useful technique in dealing with difficult-to-score items. Often, after a practice question or two, the bargaining approach can be carried out surprisingly quickly. Explain the strategy you'll use to the interviewee. (The interviewee *is* a member this two-person team!) Two bargaining strategies are suggested here, others can be devised:

a. The "Dutch Auction": Beginning with the topmost ranking ("Excellent") for the item in question, inquire whether Program performance is at that level. If the answer is "no," move to the next lower level, then the next, notching downward step by step until a level is reached that satisfies the case. Notice that if a bias is introduced by this method, it will be a bias that favors an optimistic score.

b. "Outside-In": Beginning with the highest and lowest rankings (i.e., "Excellent," and "Null"), inquire whether Program performance is at either of those levels. If it is, the appropriate score has now been identified. If not, then move to the next highest and next lowest levels. Continuing this stepwise "bracketing" will soon converge judgment to the favored score.

Mark responses to questions directly on a copy of the question sheet. Use one set of question sheets per inteviewee. Transfer scores to the computer after the interview has been completed. This will conserve interview time.
CAUTION: Again, in no case should the interviewee's name appear on the scored set of questions.

Share item-by-item results letting the interviewee know how the two of you have scored each response using the Performance Scale. This may result in useful discussion in which the interviewee alters his judgment as to the score assigned to one or more previously asked questions. Accept such changes. (See Item 17, below.)

Revisit prior items covered in earlier parts of the interview if the interviewee requests it or expresses concern. A question posed late in the interview may prompt a more thoughtful response by the interviewee than was provided in an earlier expression of viewpoint. Take advantage of such opportunities to improve the approach to the truth. Note that this does not conflict with the intent of Item 2, above.

Be objective and non-judgmental. Do not allow the interviewee to sense any personal views you may have as to program merit. Remain stolidly objective and professional in your personal demeanor, but not unfriendly. Keep in mind those things that reveal viewpoint and attitude, however subtly, e.g.: inflection of voice, selection of vocabulary, "body english," facial expression.

CLOSEOUT CAVEATS

Thank the interviewee for his time and his help in evaluating the program!

Don't divulge answers. If an interviewee asks how an item has been answered by others, named or unnamed, decline to answer. Apologetically explain that your "rule book" forbids your revealing any information about the nature of others' responses.

Do not examine cumulative data resulting from multiple interviews until all of them have been completed. Reviewing ongoing results may bias your interpretations of responses during interviews that you have not yet undertaken. If possible, assign logging of data and summarizing ongoing results to another person.

Appendix IV– System Safety Performance Level Model Workshop Attendance

5-6 June 2007

Name	Organization	Phone	5 June	6 June
Barry Hendrix	LMCO	770-494-6961	√	√
Bill Edmonds	U.S. Army Combat Readiness Center	334-255-1122	√	√
Bill Pottratz	U.S. Army AMCOM	256-313-2871	√	√
Bob Baker	APT Research	256-327-3371	√	√
Bob McAllister	WPAFB	937-257-0470	√	√
Cliff Parizo	Sikorsky	203-386-6103	√	√
Dave West	SAIC	256-971-6494	√	√
Don Swallom	U.S. Army AMCOM	256-842-8641	√	√
Donna Thompson	NGC	256-830-3392	√	√
Felisa Frazier	AMRDEC SED	256-876-1547	√	√
Gary Braman	Sikorsky	256-327-5356	√	√
Homayoon Dezfuli	NASA HQ	202-358-2174	√	√
Janet Gill	Navy	301-342-2350	√	√
Jim Schiermeyer	U.S. Army AMCOM	256-842-8623	√	√
Kerry Remp	NASA Safety Center	440-962-3188	√	√
Pat Clemens	APT Research	256-327-3707	√	√
Rhonda Barnes	APT Research	256-876-2494	√	√
Saralyn Dwyer	APT Research	256-327-3377	√	√
Tom Pfitzer	APT Research	256-327-3388	√	√
Tom Wimsutt	Raytheon	401-842-3615	√	√
Willie Fitzpatrick	AMRDEC SED	256-876-9945	√	√

Appendix V – System Safety Performance Level Model Workshop Minutes