Safety Process Measurement – A Review

Report Authors
Paul Caseley
DSTL/CP06715 V1

6 May, 2003

Paul Caseley
N133
DSTL Malvern
St. Andrews Road
Malvern
WR14 3PS
UK

THIS DOCUMENT IS THE PROPERTY OF HER BRITANNIC MAJESTY'S GOVERNMENT, and is issued for the information of such persons only as need to know its contents in the course of their official duties. Any person finding this document should hand it to a British Forces unit or to a police station for safe return to the Chief Security Officer, DEFENCE SCIENCE AND TECHNOLOGY LABORATORY, Porton Down, Wiltshire SP4 OJQ, with particulars of how and where found. THE UNAUTHORISED RETENTION OR DESTRUCTION OF THE DOCUMENT IS AN OFFENCE UNDER THE OFFICIAL SECRETS ACTS OF 1911-1989. (When released to persons outside Government service, this document is issued on a personal basis and the recipient to whom it is entrusted in confidence within the provisions of the Official Secrets Acts 1911-1989, is personally responsible for its safe custody and for seeing that its contents are disclosed only to authorised persons.)

Dstl is part of the
Ministry of Defence

Number of Pages in Report 40

Release Conditions
This document has been prepared for MOD and, unless indicated, may be used and circulated in accordance with the conditions of the Order under which it was supplied.

It may not be used or copied for any non-Governmental or commercial purpose without the written agreement of Dstl.

© Crown copyright 2003
Defence Science and Technology Laboratory UK

Approval for wider use or release must be sought from:

Intellectual Property Department
Defence Science and Technology Laboratory
Porton Down, Salisbury, Wiltshire, SP4 0JQ.

Table of Contents

List of Figures

List of Tables

1 Introduction

For a system to operate safely operators, suppliers and developers must ensure that all the safety aspects of the system have been considered and assessed as safe. Safety processes are used to assess and measure the safety risk of a system, its operations and procedures. For example a hazard analysis HAZOP study is a safety process that can identify hazards (safety products) for a particular system. However, there is clearly an issue as to what degree of confidence should be placed in the products of safety processes. This report reviews methods that can measure management, effectiveness and quality of safety processes, not how the safety of a system is measured and justified. Much of the context of the report is from a United Kingdom Ministry Of Defence (MOD) safety perspective but much of its content is equally applicable to other industry safety domains.

The investigation of safety process measurement includes:

- measurement of processes, especially current practice for software;
- terminology for safety to help bound and identify entities and attributes related to safety processes;
- organisational and competency measurements related to safety processes;
- products of safety processes;
- frameworks for measurement and
- existing safety processes measurements.

1.1 Background, Consequences and Motivation

Safety is an important issue to the public, suppliers and procurers. In recent years, the United Kingdom Ministry Of Defence (MOD) has had to respond to government legislation by improving its equipment safety management systems and accepting increased responsibility for equipment safety. The author estimates that safety costs anywhere between 1% and 15% of procurement and support[1] depending on the equipment and its role.

For the MOD, some safety issues are more complicated due to the unique problems faced by military systems. These systems often push acceptable safety boundaries in their need to outperform and defeat an enemy. To ensure safety assurance and still meet operational requirements, MOD has set challenging equipment safety Defence Standards (DSs) ([1], [2], [3] and [4]) based on the policy of making the safety risk "As Low As Reasonably Practicable" (the principles of ALARP are discussed in HSE [5]).

[1] The range of these figures is speculative. However, for one MOD risk class C project, in assessment phase, an approximate figure of 1.3% has been calculated. Unfortunately the details in deriving this figure are constrained by commercial issues.

The ALARP equipment safety policy is implemented by MOD Safety Management Systems that are "based on a Safety Case approach that addresses all aspects of through life safety for the equipment", Joint Service Publication (JSP) 454, [6]. The equipment Safety Case itself depends on all the necessary safety activities (processes) that provide the safety evidence and safety assurance arguments.

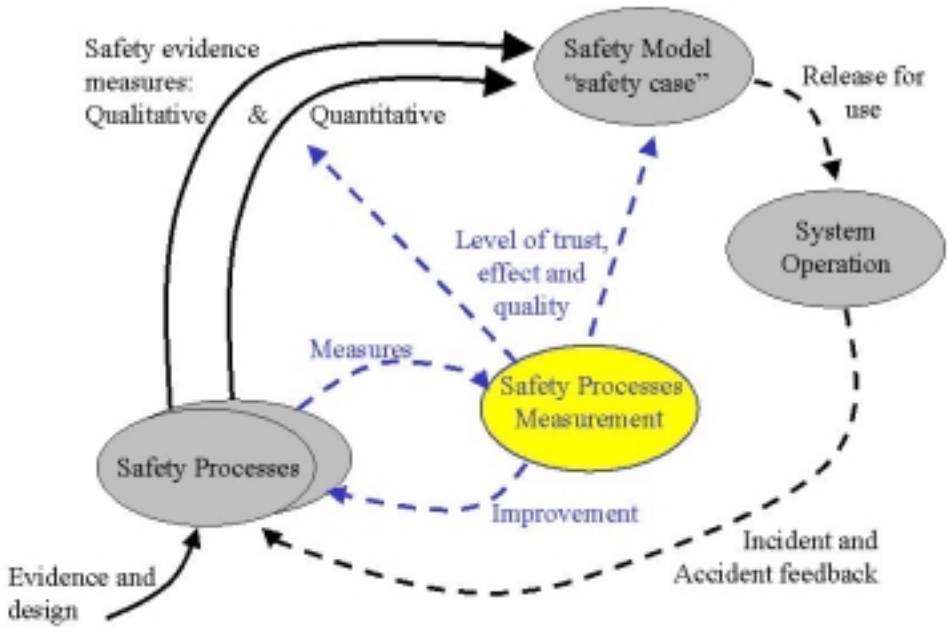

Figure 1-1 Measurement of safety and safety processes

This report reviews possible mechanisms for measuring safety processes that generate the safety evidence and the possibility of combining such mechanisms into a general safety process measurement framework. From such a framework it may be possible to establish the degree of trust that should be placed in safety processes and therefore the degree of trust we should place in the safety evidence and argument analysis results. The principle is illustrated in Figure 1-1. The report format is as follows:

- Section 2, an introduction to measurement for safety specialists
- Section 3, an introduction to safety issues for measurement specialists
- Section 4, measurement techniques used for safety processes measurement
- Section 5 measurement techniques that could be adapted to safety processes measurement
- Section 6 summary and conclusions

2 An Overview of the Theories and Principles of Measurement

This section is intended to give a background to measurement principles. It is only an overview and is aimed at engineers and managers not familiar with measurement theory who may need more contextual information to appreciate the issues of measuring safety processes. Those familiar with measurement should skip to section 3.

2.1 Fundamentals of Measurement

The importance of measurement should not be underestimated. It is one of the foundations of civilisation and impacts many of our daily activities.

A description or definition of measurement must include its diverse subjects such as: comparing weather, dimensions of objects and economic events. An attempt to informally define measurement by Fenton and Pfleeger [7] describes measurement in abstract and real world terms with predicates:

"Measurement is the process by which numbers or symbols are assigned to attributes of entities in the real world in such a way as to describe them according to clearly defined rules"

From above it is relatively easy to map entities to real world objects or events:

- Objects: person, wardrobe, liquid, etc;
- Events: test, journey, accident, etc;

Events can also be mapped to the products of processes or the processes themselves. The products of the test phase in a software process are the results, which are usually formally documented. Similarly the accident sequences related to hazard environments can be documented in the Preliminary Hazard Analysis (PHA) for a system.

An important aspect of measuring an entity is its attributes. These are characteristics or manifestations that add to the description of the entity. Typical attributes are: length, mass, and volume.

Attributes obey rules, which put the measurement into a context. The rules are usually embodied in numbers and symbols, that is: an attribute length can be a number, e.g. 5m; an attribute risk could be a symbol, e.g. risk class A. The rules for length measurement and length are precisely defined, in this case the metric system.

For safety some entities and attributes are difficult to define and measure. Typical of these are Hazards, which can be objects or situations. The lack of a universally acceptable definition of a Hazard highlights the predicament – Meulen [8] has identified eleven Hazard definitions. The confusion caused by terminology in the safety domain has been raised by many, including Lees [9], who states:

"Unfortunately there is at present no accepted terminology in this field"

2.1.1 The Representational Theory of Measurement

Fenton's and Pfleeger's natural language definition is an overview of the modern form of measurement referred to as the representation theorem or the representational form of measurement. It provides sufficient understanding for most measurement practitioners, however, formal measurement theory is a significant mathematical topic. The formal theory requires three steps to be defined:

1. A description of an empirical relation system – a definition of the attributes and entities.

2. A representational theorem – the rules that apply to the attributes and entities.

3. A uniqueness condition – definition of the scale used for measurement.

When considering new measures the engineer should be aware of these three requirements. An overview of the theory can be found in Finkelstein and Leaning [10].

2.1.2 Direct Measures

Direct measurement, sometimes referred to as a base measure, is the simplest form of measurement structure where only a single attribute, property or manifestation of an entity or process is being measured. Even when using the simplest forms of measurement you should have rules and a scale for the comparison with the real world for the measurement to be meaningful. An example is an everyday measurement such as Volume.

2.1.3 Scales

When discussing scales they are often classed as four major types: Nominal, Ordinal, Interval and Rational. They form a hierarchy and are described in more detail by Stevens [11] the lowest being Nominal the highest Rational. The higher scales allow greater mathematical manipulations. It is therefore important to understand the limitations of particular scales.

Nominal scales are useful to allocate measurement to clearly defined and separate entities or groups of entities that cannot be separated from their measurement label. For example, colours could be allocated a nominal scale as red is separate from yellow, however, there may be issues separating differing shades of orange colour.

Ordinal measurement systems display a weak ordering that allows comparison but no addition or subtraction. Shades could be a type of ordinal measurement, e.g. light, medium, dark. However, an element of subjectivity enters into such a measurement scheme, although it may be possible to discriminate between large differences of shades the point at which a

shade moves from light to medium could be disputed. For ordinal measurement to work, correct comparison must be possible.

Accident severity categories

Category	Definition
Catastrophic	Multiple deaths
Critical	A single death; and/or multiple severe injuries or severe occupational illnesses
Marginal	A single severe injury or occupational illness; and/or multiple minor injuries or minor occupational illnesses
Negligible	At most a single minor injury or minor occupational illness

Table 2-1 Accident Severity table from DS 56

The accident Severity[2] table above, extracted from DS 00-56, also demonstrates clearly the usefulness of a simple ordinal measurement. There is no real advantage to allocating numbers to severity, other than making the weak ordering more obvious, as the addition of two severity may not change the overall severity or even increase it, e.g. a combined fire and flood accident may produce a reduced combined accident severity.

The Interval scale is extensively used allowing mathematical functions such as addition and subtraction. The intervals can be regarded as equal, but the context of the interval has to be correctly used to be meaningful. For example, a measurement of a library could be the number of books it holds. However, a particular library may hold only one type of book and despite having more books, it may have less information than other libraries. With interval scales it is possible to express sensible difference statements but not meaningful ratio statements. For example: Joe has an IQ of 120 points and Bill has an IQ of 60, the difference is 60 points but Joe cannot be said to be twice as intelligent as Bill as the zero point for IQ measurement is purely arbitrary. From the previous example of counting books, a natural zero, i.e. no books, is implied and this is the distinguishing feature of the most useful scale, the ratio scale.

The ratio scale has an absolute zero point that enables meaningful ratios to be made, e.g. Joe's Library has twice as many books as Bill's. Counting entities nearly always implies a ratio scale but not necessarily a useful measurement. Counting hazards on aircraft systems exhibits some of the characteristics of a ratio measurement system. It is possible to count the number of hazards on two identical aircraft. Adding systems to the aircraft may change the number of hazards and taking away these systems also change the number of hazards by an equal amount. Despite this, it does not necessarily mean that the overall hazard is increased or decreased when a system is added. If one of the aircraft added weather radar this will increase the number of hazards but may decrease the overall risk of an accident, whereas the addition of a cargo hold for carrying dangerous chemicals will also add hazard but probably increase the overall risk of an accident. Thus the counting of hazards is monotonic but it

[2] Severity is the term used in DS 00-56 and more general term is consequence.

does not mean the overall risk is monotonic. This is one of the reasons why counting hazards as a safety measurement is only of limited use.

2.1.4 In-Direct, Derived Measurement

A derived or In-Direct measurement is composed of two or more independent attributes. These attributes should have direct measures or have a derived measure based on direct measures. A safety example of a derived measurement is table 5 of DS 00-56 part 1, which uses an ordinal severity measure and a probability measure (interval) to derive accident risk. This derived measurement has a subjective element because its base measures and their interpretation varies between systems - DS 00-56 part 2, table 2, has a similar table based on the Nuclear domain and derives different accident risks. This example illustrates two points:

- Subjective measurements, no matter how useful, depend on the environment in which they are made. Their representation condition (rules) may not be formally defined;
- The best scale measurement for a derived measurement will be the weakest of the input sub-attributes scales.

The latter point is important, as the preferred derived or in-direct measure is composed from independent component attributes that have ratio scales. Thus the derived measurement scale will be also ratio and will have the increased statistical analysis options inherent in the higher order scale.

2.1.5 Practical Measurement Models, Problems and Solutions

A measurement requirement is usually driven by a business or scientific need for information in order to understand a process, phenomenon or entity. For successful measurement McGarry et al [12] details important planning, implementation, data collection and analysis activities that lead to the formation of a measurement programme. McGarry et al describes the basic tasks of a measurement programme as follows:

- Information Need – identifying the scientific, engineering or business requirements;
- Measurement Concept – Creating the representation theorems and scales (key stages of formal measurement Fenton & Pfleeger);
- Measurement Construct – Combining the measures into information that meets the need;
- Measurement Procedure – the mechanics of applying the measurement in a consistent manner;
- Measurement Plan – a combination of the above in a documented form.

It is difficult to choose a useful measurement in a domain that has been previously ignored, or is very complex, because of the vast amount of different measurement possibilities. Park et al [13] and Florac et al [14] highlight this problem of choice and complexity and suggest a goal-oriented approach (this is assessed later).

Another solution, is to build new measurement constructs onto an existing measurement model - an example of how object oriented software can be measured using a measurement information model based on traditional software measurement techniques is described by Card et al [15].

2.1.6 Summary - Theories and Principles of Measurement

A measurement system has a set of rules that map symbols to attributes of something that needs to be measured. The symbols usually take the form of numbers and the rules can be formally defined as representational theorems. The uniqueness of measurement also needs to be defined and dependent on chosen measurement scale. There are four general scale types: Nominal, Ordinal, Interval and Ratio. They form a hierarchy of scales the latter scale types are the most powerful and allow increased mathematical and statistical manipulation.

Measurement is driven by a need, which should be defined before the creation of representations or scales. Often these needs are met by combining or building on existing measurement models.

3 Issues with Measurement of Safety Processes

This section investigates and highlights potential problems to measuring safety processes. It is aimed at measurement specialists who need additional context information in order to understand the problems in measuring safety processes. Those familiar with safety should skip to section 4.

3.1 Definitions of Process

Like many terms used in engineering the term *process* is overloaded and potentially ambiguous. A prescriptive definition is attributed to Gabriel Pall by Florac et al [14]:

"A process can be defined as the logical organization of people, materials, energy, equipment and procedures into work activities designed to produce a specified end result."

The above definition identifies a number of potential attributes and objects that could be measured, i.e. people, material, etc. It also implies that a process should include organised people resources, which may not be the case for every process, e.g. automatic archiving of data. Two other simpler and more general definitions are:

"An organized set of activities performed for a given purpose." Mil Std 498 [16]

"A set of interrelated activities, which transform inputs into outputs" ISO/IEC TR 15504 part 9 [17]

Pall's definition has some advantages in that it identifies interfaces and relationships on processes. Fenton & Pfleeger [7] have recognised this in that they identify three classes of software measurement, *Processes*, *Products* (that result from process activities) and *Resources* (entities required by processes). Their software classification scheme clearly shows a relationship between products and resources to that of processes. This is a principle that could be transferred to safety processes. Other classification schemes have also been proposed, e.g. around artefacts, activities and agents, Armitage et al [18].

3.2 Safety Processes

The term "safety processes" is not defined but encompasses all safety activities and techniques that produce products that in turn support the Safety Management System. Typically, a safety process is instigated by an event, such as requirement for hazard identification, this is illustrated in Figure 3-1 extracted from Caseley, Clark and Powell [19].

A process produces some outputs, e.g. a design assumption or an analysis report. In Figure 3-1, a hazard list and a set of accident scenarios might be produced.

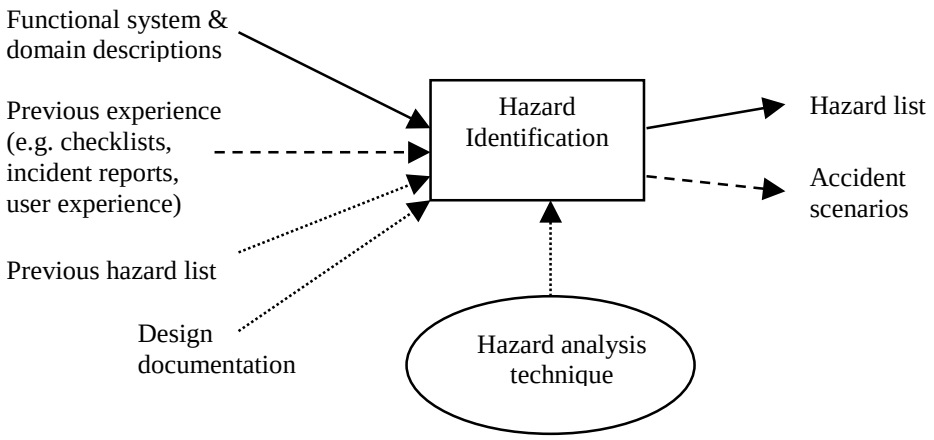

Figure 3-1 Inputs, output products and resources for Hazard Analysis

3.3 Safety Definitions

As discussed in section 2.1, there is not one standard set of descriptions of safety concepts and terminology in the safety field and many terms are overloaded with multiple definitions. The American Institute of Chemical Engineers (AIChE) [21] identified eight differing synonyms for their description of Hazard Evaluation. Even simple atomic attributes such as the term accident have differing definitions, e.g.

Accident –

Unintended event or sequence of events leading to harm. ISS [20].

An unplanned event or sequence of events that results in undesirable consequences. AIChE [21].

An incident with specific safety consequences or impacts. Also AIChE [21].

An unintended event or sequence of events leading to death, injury, environmental or material damage. DS-56 [1].

Mishap: An unplanned event or series of events resulting in death, injury, occupational illness, or damage to or loss of equipment or property, or damage to the environment. An accident. Mil Std 882C [22].

An accident is an undesired and unplanned (but not necessarily unexpected) event that results in (at least) a specified level of loss. Leveson [24]

Much of the terminology overload can be traced to differing industry domains and the differing use within the engineering disciplines, Tribble [23]. However, in many cases the synonyms are clear and traceable. The safety terminology of a project may have to be defined in order to ease identification of attributes and implementation of the a safety process measurement programme.

3.3.1 Safety Processes – Boundaries

Figure 3-1 illustrates inputs and issues that may trigger Hazard Identification and potential outputs of the process. The process Hazard Identification may use a number of sub safety processes such as Fault Tree Analysis (FTA). Safety processes are not limited to the design and development they are also used in the operational and disposal phases. Furthermore, their analysis also covers aspects of occupational and environmental safety. Thus the scope of safety processes is large and this reports review considers many diverse domains in order to understand the boundaries of safety process measurement. Any safety process measurement programme will have to clearly define the boundaries of the measurement process.

3.3.2 Safety Lifecycles and Techniques

Defining the temporal order of safety processes and identifying which are the most applicable safety processes for a development phase is also a challenge.

A key safety process in DS 00-56 [1] is hazard analysis, which is performed throughout the phases of a project from the concept stages, right through to decommission and disposal. It is nominally split into three phase parts: Preliminary Hazard Listing (PHL), PHA and System Hazard Analysis (SHA). The PHL, PHA and SHA are a cycle of activities that iterates through identification evaluation control. Projects using DS 00-56 are urged to start hazard identification and refinement at the outset of a project, however, the mapping of hazard analysis phases to development and operational lifecycles has a considerable degree of flexibility.

Standards, such as ARP 4754, have attempted to define a lifecycle of specific safety analysis techniques in a set temporal order. An example is Figure 3-2, extracted from Murdoch, McDermid, and Wilkinson [27], which is a variation on the ARP 4754 [26] model and clearly defines the order of safety processes with respect to a typical system development lifecycle. A more prescriptive example of temporal ordering of safety activities in the lifecycle is illustrated in figure 2 of BS IEC 61508 part 1 [25], which defines 16 phases to be performed to achieve safety. Specific allocation of safety processes to a design phase has advantages when considering process measurement.

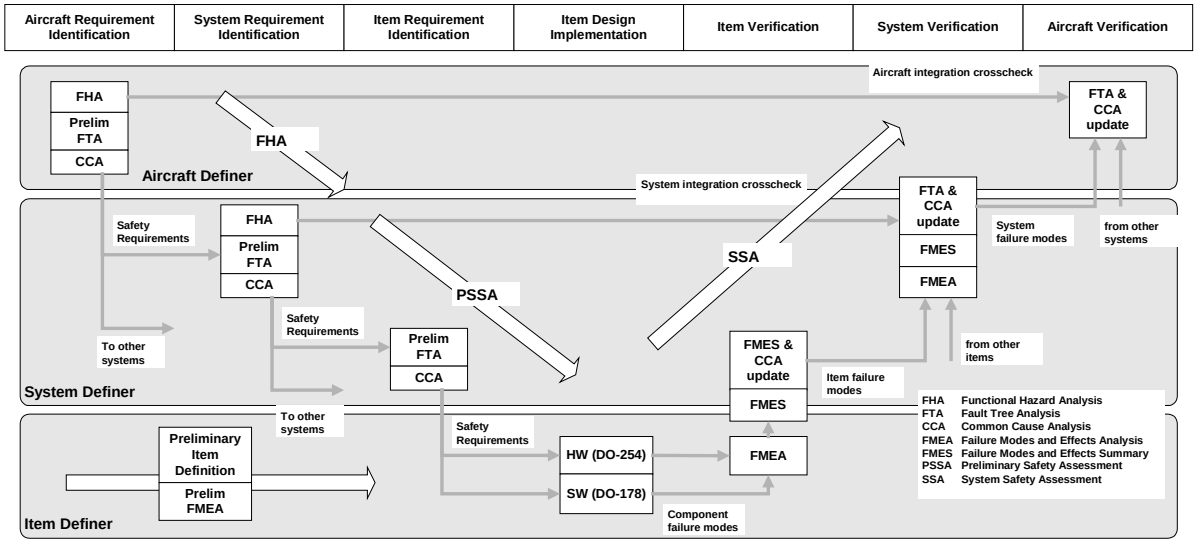

Figure 3-2 Safety processes mapped to safety and system development lifecycle

Safety lifecycle models, similar to that illustrated in Figure 3-2, allow measurement comparisons to be made against development phases; e.g. how much effect the preliminary FTA has on the System Requirements Identification development phase. Such comparisons would be useful, but may not be practical, as there is no point in comparing data against theoretical lifecycles that are never actually used within industry.

This is could be a key question for a safety process measurement program. That is, are the safety processes being applied at the correct times so that they influence and drive a safe design?

3.4 Summary - Issues with Measurement of Safety Processes

Processes can be identified and measured. They produce products and depend on resources, both of which can also be measured.

The definition of safety terms is diverse and many terms are overloaded. To ease the identification of measurable entities and attributes within safety processes the terminology of the measured safety processes may need to be defined. The scope of measuring safety processes is very large and the boundaries of a safety process measurement programme may also need to be clearly defined.

The Defence Standards are flexible and do not rigorously prescribe which safety processes should be used or when, with the exception of HAZOP. Thus, there is no obvious baseline model for a safety lifecycle that maps particular safety processes to system design phases. The comparison between objects will be problematical, thus, pragmatic solutions may be the only option for a project.

4 Possible Measurement Strategies for Safety Processes

This section surveys current and past work directed at investigating and measuring safety processes. It is possible that the survey is incomplete and the author would appreciate readers forwarding relevant information to the contact address on the front cover page.

4.1 Guidance on Measuring Effort

One of the few references found containing safety processes effort estimates for use by safety is the Guidelines for Hazard Evaluation Procedures [21]. This book is the product of the Hazard Evaluation Procedures sub-committee at the Centre for Chemical Process Safety established by the AIChE. It contains the experience of many safety specialists based in the chemical domain. The work deals with qualitative hazard evaluation which is a subset of safety processes typically carried out during PHA and SHA in MOD projects. The Guidelines document the effort estimations for twelve safety processes but the authors caveat these by stating:

"However, estimating the time and effort needed to apply a particular HE technique is more art than science, because the actual time to perform a study is influenced by many factors – some of which are not quantifiable"

The Guidelines particularly express reservations on estimating the complexity and size of the system, nevertheless they provide rough estimates for preparation, evaluation and documentation of effort based in on the estimated size/complexity of the system or process for all twelve safety processes. Some of the figures are summarised in Table 4-1 and Table 4-2.

Technique	Complexity/Size	Preparation	Model Cons-truction	Qualitative Evaluation	Document-ation	Total (Hrs)
Fault Tree Analysis (FTA)	Simple/Small System	8 to 24	24 to 48	16 to 48	24 to 40	120
	Complex/Large Process	32 to 48	80 to 120	40 to 160	120 to 160	440
Human Reliability Analysis	Simple/Small System	4 to 8	8 to 24	8 to 16	24 to 40	48
	Complex/Large Process	8 to 24	40 to 80	40 to 80	40 to 120	280

Table 4-1 Some examples of model based Hazard Evaluation effort estimations

Technique	Complexity/Size	Preparation	Evaluvation	Document-ation	Total (Hrs)
What-If Analysis	Simple/Small System	4 to 8	4 to 8	8 to 12	28
	Complex/Large Process	8 to 24	24 to 40	40 to 120	184
HAZOP	Simple/Small System	8 to 12	8 to 24	16 to 48	84
	Complex/Large Process	16 to 32	40 to 120	80 to 240	392
FMEA	Simple/Small System	2 to 6	8 to 24	8 to 24	54
	Complex/Large Process	8 to 24	40 to 120	80 to 160	304

Table 4-2 Process based Hazard Evaluation effort estimations

4.2 Process Engineering Language

Process Engineering Language (PEL) is a system performance measurement technique (based on effort) which originated from a method of expressing a Generic Work Breakdown Structure at BAe Systems. Internal research and co-operation with close aerospace partners led to the development of a language based around the actual activities performed by software engineers.

Clark and Powell [28] describe PEL as a finite state language that expresses technical process information in a consistent manner. Using a common language to express software activities Powell [30], and also Clark and Powell were able to compare Rolls-Royce and BAe software projects. PEL has rules of syntax based on a simple template of actions, phases, products and representations. The template allows descriptive sentences to be constructed. The granularity of the descriptive sentences permits subtle inferences and interpretations to be made. Clark and Morris [29] describe how a PEL sentence is constructed. A typical example uses simple joining phrases to make the sentence parse as a natural language statement, e.g.

"*In the* <PHASE>, <PROCESS> *the* <PRODUCT>". When applied using the standard lists available to the engineer translates to "*In the* Preliminary Design *Phase*, Estimate *the* FUG_FUM_DATA"

PEL requires a lexicon of engineering terms to be defined but allows synonyms and comparisons. A variation of the software lexicon is being developed to measure safety effort as part of a MOD Research programme. The initial concepts were explored by Caseley, Clark and Powel [19] and an outline of a generic safety lexicon with discussion on some of issues regarding its use can be found in a MOD research report [31]. An addition in SPEL over the original PEL, is the introduction of an extra product dimension that represents the safety products, e.g. fault trees, event trees, etc. The representation is a transformation of the design into a safety product. Using the Table 4-3 can lead to construction of safety related statements with associated effort, e.g.

incident, investigation, produce, accident report, Aircraft, Guidance, Control, Actuator : 30 hours

This can then be interpreted as a SPEL sentence by adding additional joining words:

"For the air incident investigation, producing an accident report on the Aircraft Guidance Control Actuator took 30 hours."

Safety PEL dimensions				Project PEL dimensions			
Phase/Event	Process	Action	Representation	Platform	System	Sub System	Unit
Requirements In-Service Incident DRACAS …	Hazard- Identification Investigation …	Produce Review …	Fault Tree Safety Case Incident Report …	Aircraft Ship Hospital ….	Guidance Engine Navigation ….	Control Display ….	Actuator Pipe Code …

Table 4-3 Basic Safety PEL structure with representation, Caseley, Clark and Powell

4.3 Cost Based Analysis and Return on Investment

Investigations into the benefit of differing safety processes as a cost benefit process has been investigated by Tribble [23] and extended to organisations by Jervis and Collins [32]. Both rely on an analytic hierarchy process (Satty [33]) for determining a relative numerical ranking against alternatives.

Jervis and Collins conducted a management survey to assess the Return On Investment (ROI) of safety programs on a number of military facilities that maintained the US Occupational Safety and Health Administration's Voluntary Protection Program (VPP) framework. The VPP framework is applied to Federal sites and consists of seven major safety related elements. The survey assessed six the "managerial-type safety program elements" the exception being the accident experience element.

In the survey managers were asked to subjectively compare each of the safety elements against each other and rank their relative importance in the two areas of relative resources and relative benefit. This qualitative comparison was transformed into quantitative data for analysis of cost benefit issues. Their work reveals that the VPP element "hazard prevention and control" provided the greatest benefit where they had intuitively expected that the second ranked "management leadership and employee involvement" to provide the greatest ROI for the safety programs. Surprisingly the survey revealed that "safety and health training" ranked a very poor fifth. This method is best suited to organisational issues related to occupational safety and not functional safety but highlights the advantages of measurement within a known context/framework.

Tribble's survey examined the relative benefits of "Return" and "Investment" for the differing safety assurance processes for safety critical air systems. The survey suffered from poor returns so the analysis should be regarded as statistically invalid. Nevertheless the results provide some useful indications of what engineers and managers saw as the most valuable safety processes. The results indicated that certification is the most costly in terms of investment and return, and that design verification ranked consistently high for ROI. Furthermore, Tribble asserts that

"The high ROI associated with design verification is an indication that more attention should be paid to rigorous proof techniques like formal methods."

This is perhaps a justification for the direction of measurement priorities for a particular project, that of safety process associated with verification of software.

4.4 Measuring Hazard Identification

Empirical evaluation work of safety analysis techniques within industrial applications by Suokas [34] compared a number of techniques used for hazard identification, including the MOD preferred method - HAZOP study. Suokas attempted to derive measurements for reliability, coverage and validity of safety analysis techniques (safety processes).

Suokas included a number of case studies in his work and the research clearly demonstrated some of the effects of competency of analysts, diverse analysis techniques and complexity of systems. The case studies showed that:

- Some hazards cannot be easily identified by some safety analysis techniques. This demonstrates that diverse techniques are essential for complete hazard identification. The work showed that it maybe possible to set the boundaries/scope of a particular analysis technique. Other techniques could be used to complement and gain a more complete coverage of Hazards. Although not fully covered in the work, it is likely that some techniques are more effective in some engineering domains, e.g. HAZOP may be more effective in the chemical process domain than in software[3].
- Competency and experience can affect the analysis, although not always as predicted – in one case the most experienced operator found the least hazards.
- The nature of the system probably influences the results – in one case over 40% of hazards of one sub-system were not identified.

For the most part, the studies depended on multiple teams using diverse techniques as well as using accident and incident data as a feedback mechanism to confirm existing hazards or identify new ones. Suokas also identified that the quality of the data about the system or work practice is also a factor in determining hazards. Suokas proposed the following formulae for interanalyst reliability, coverage and validity:

$$Reliability = \frac{\text{Number of hazards identified by a test person or Team}}{\text{Total Number of hazards identified in the experiment}}$$

$$Coverage = \frac{\text{Number of hazards identified by Method examined}}{\text{Total Number of hazards identified in the experiment (and belonging to the search pattern of the method examined)}}$$

$$Validity = \frac{\text{Number of hazards identified by the method examined and in accident reports}}{\text{Total Number of hazards identified in accident reports}}$$

Suokas's results showed that measured internalyst reliability reached 100% for identified hazards when using HAZOP in only one of the three experimental studies. This reliability dropped to as low as 50% when factors such as deviations (events leading to possible accidents) and determining factors (constant safety pitfalls) were taken into account. Corresponding coverage and validity for HAZOP were measured at 89% and 71% respectively.

[3] This is perhaps not the best example as there are specialist software HAZOP techniques, unfortunately it is difficult to compare.

These figures are only indicative because of the limits of the experiments and the complexity of the systems examined. Applying these measurements to a project may be difficult and costly, as they require:

- the use of multiple teams of analysts;
- the application of multiple techniques for comparisons and
- a system history with accident and incident information.

Suokas's work raises important points with regards to diversity and effect of differing resources (techniques and personnel). However, this work does not seem to have been widely referenced or accepted in industry. This would suggest Suokas's formulae are not as successful as his research indicated.

4.5 Assessing the Quality of Safety Processes using Checklists

Quality Assessment of Safety Analysis (QUASA) investigated and developed at VTT by Rouhianinen [35], is a technique intended to support quality management of safety analysis. It is based on the hypothesis that quality depends on the adequacy of the safety analysis process and that the adequacy is the result of planning, organisation and execution of the analysis.

QUASA is a method based on a checklist for particular lifecycle phases and safety processes. The checklist supports the quality assessment of the safety analysis and is intended to be a tool for quality control. The checklist statements have no ranking, as their importance is said to vary according to the subject analysed. The method described by Rouhianinen has been applied to compare two safety analyses (HAZOP). The results highlighted a number of problems:

- Risk Assessment – hazards were identified but their risk was incorrectly classified. The examples cited indicated that 14% of the hazards identified in a HAZOP study classed as negligible risk subsequently led to a recorded incident.
- Omission – The HAZOP study missed the hazards. Operating incidents showed that 2 studies missed at least 16% of the hazards, further confirming the finding of Suokas.
- Wrong scope – the HAZOP study did not consider the whole system or included elements of the environment that were not relevant. This reflects that HAZOP, the safety process studied in this research, is not sufficient to cover all aspects of a system's safety. Again this further supports Suokas [34] who asserts diversity of analysis is essential.

Overall application of QUASA to the two safety analyses suggested that the HAZOP studies prevented about 65% of hazards.

Auditing using checklists is a common method of measuring a process or part of a safety management system. Extensive guidance on what auditing is capable of assessing for safety and environment can be found in essays collected by Harrison [36]. Auditing is best done

against a form prescriptive standard so that compliance can be assessed. Auditing is based on professional judgements with the output not a measurement but a report containing a "true opinion", pp36, Harrison.

4.6 Selecting and comparing Safety Processes

Efforts have been made to compare safety analysis techniques by Rouvroye and Bleik [37]. Their analysis uses a series of qualitative and quantitative checks grouped in three viewpoints, that of: system information needed (design input to the process), Actions performed (effects on the process on the design) and overall safety output (obtained results). The comparison was limited to a sub-set of "popular" techniques from those defined in IEC 61508, the choice is not justified and omitted ETA and CCA but included two versions of Markov analysis. The authors claim that Enhanced Markov analysis should be preferred because this technique covers the most aspects relevant for the quantification of safety.

A comparison of hazard evaluation techniques with selection advice can be found in AIChE Guidelines for Hazard Evaluation Procedures, chapter 5. The guidelines make a strong argument based around the unique strengths and weaknesses of each technique and the availability of local knowledge and expertise. No single technique is highlighted as the best or graded against the others.

4.7 Summary - Possible Measurement Strategies for Safety Processes

The measurement strategies surveyed are few due to the lack of research effort in this area and do not appear to have been widely accepted in Industry. Even basic estimates for safety process effort are very crude and variable[4].

PEL shows promise in that it can overcome the overloading of terms used in safety and measure efforts very accurately. This should allow more accurate comparisons between projects something that PEL projects have successfully achieved in the software measurement.

Return on Investment research has shown that a known context or framework has advantages when making measurement comparisons. ROI safety research has also indicated, albeit not conclusively, that research in the verification processes for software are important.

[4] As previously highlighted, MOD research has some basic effort figures to compare against those provided by the AIChE but they cannot be released due to the project's sensitive nature.

5 Other Methods and Relevant Work

This section investigates methods that could:

- identify safety process measurements;
- be adapted for safety process measurement or
- aid in the measurement of safety processes.

5.1 Goal Based

The selection of measurement attributes for safety processes is an important step in building a safety measurement framework. A widely used top down method of selecting metrics is the Goal-Question-Metric (GQM) process [38]. This method concentrates on business goals and questions that should be asked to meet them. Having set a business goal the analysts ask questions such as, "what is it that I need to know?", instead of "what measurement should I take?".

Experience on the application of GQM has lead some practical problems which has lead to criticisms by Bache and Neil [39] and Wearing [40], particularly in the following areas:

- Top down approaches often miss practical solutions when detailed analysis reveals simple solutions.
- GQM can suggest attributes that are very difficult to gather and may only give transitory short term benefits.
- Many of the goals are unachievable – this is usually the case for large projects where the customer requirements are very challenging.

Nevertheless, GQM can be an effective brainstorming technique for eliciting and documenting measurement objects, and efforts have been made using GQM to scope this research some of which can be found in MOD research [31].

Much of the criticisms of GQM have been addressed by Park et al [41] where they introduced a slight variation called GQ(I)M where the I introduces the term "Indicator". Their process consists of three precepts which map to ten steps. It is effectively a design process for developing measurements that introduces elements of common sense and practicality. As with all design processes, there is feedback as you progress through the ten steps. The process is considered top down but much bottom-up work is required as the design process evolves toward a set of measurements that are useful.

5.2 Practical Software and Systems Measurement

Practical Software and System Measurement (PSM) is a Department of Defence (DoD) initiative with wide support in the US defence industry. Smith [42] describes PSM as "a systematic, flexible, and objective process for analysing software and systems development project issues, risks and financial management". The initiative is intended to link the DoD's smart acquisition to measurement schemes and process improvement methods such as ISO 9000 [43] and the Capability Maturity Model (CMM) [44] sponsored by the Software Engineering Institute (SEI) who are one of the more influential contributors to PSM. The PSM viewpoint concentrates on measuring processes, i.e. "using measurement to manage and improve software processes", PSM Guidebook [12].

PSM Version 3.1a [45] is a software process measurement programme but is currently being widened to include system processes in the draft PSM 4.0b [46]. The draft version does not explicitly cover safety but it does indicate a need for safety related measurements. There is some effort in part 3 of PSM 4.0b "Measurement Selection and Specification Tables" to identify safety related attributes associated with system development and system operations. A tertiary tabular hierarchy is used to compartmentalise measurements into common *issues* and then measurement *categories* which contain *measurement* data items (referred to as I-C-M tables). Table 5-1 is all the safety I-C-M data extracted into one table and indicates how little attention has, so far, been paid to measuring safety processes in PSM.

The application of some the principles of PSM for safety has been identified by Smith [42]. In this paper, three measures are proposed for to evaluate operational safety as defined by the AFPD 63-12 [47]. The measures were:

- Problem reports: Critical failures of a particular system over 10 years.
- Failures by Cause: Total accidents per year compared to accidents attributed to a separate process, the example used was maintenance.
- Experience Level: The US Air Force allocates skill levels to its personnel and Smith tried to relate the accident measures to the changing skill levels in the Air Force.

There is a large amount of data available in the public domain on PSM and the PSM community has had, to date, six annual conferences. The approach and philosophy of PSM is consistent and uses terms defined in international measurement standards such as the ISO/IEC 15939 [48]. From a safety process measurement perspective, PSM is attractive because it has a system approach that includes safety. PSM considers measurement a key function in process management and improvement. It emphasises that good measurement requires planning.

Common Issue Area	Measurement Category	Measures	Data Items; Attributes
Product Quality	Efficiency	Utilisation*	Maximum capacity of resource, Maximum amount of resource established as design limit, maximum amount of resource established as performance limit, Date/time of measurement, Amount of resources used Resource type, Increment, State or Mode Operational Profile, Function , task or operation measured, Test sequence
	Usability*	Operator Errors	Time period over which task was performed, Number of operators errors; Task identifier, Increment, User interface device, Priority, Test sequence, Category of operator errors, Operations document identifier
	Dependability – Reliability*	Fault Tolerance*	Number of single point failures, Number of identified failure modes, Number of identified failure modes with fault-tolerant design protection; Failure mode, Failure effect, Redundancy level, Type of Fault
Customer Satisfaction	Customer Support	Request for Support*	Number of requests , Number of reported defects; Increment, Priority (safety hazard, critical impact, minor), Type of support requested, Request mechanism, Non support resolution (request beyond support agreement), Status code (open, closed) Customer or originator of request, Activity when problem was discovered.
		Support Time*	Number of requests received, Average response time, Maximum response time, Average time to resolve, Maximum time to resolve Type of maintenance required, Increment, Priority (safety hazard, critical impact, minor), Non support resolution (request beyond support agreement), Customer or originator of request, Request mechanism.

Table 5-1 extracted from PSM 4.0b, safety related C-M-I table data

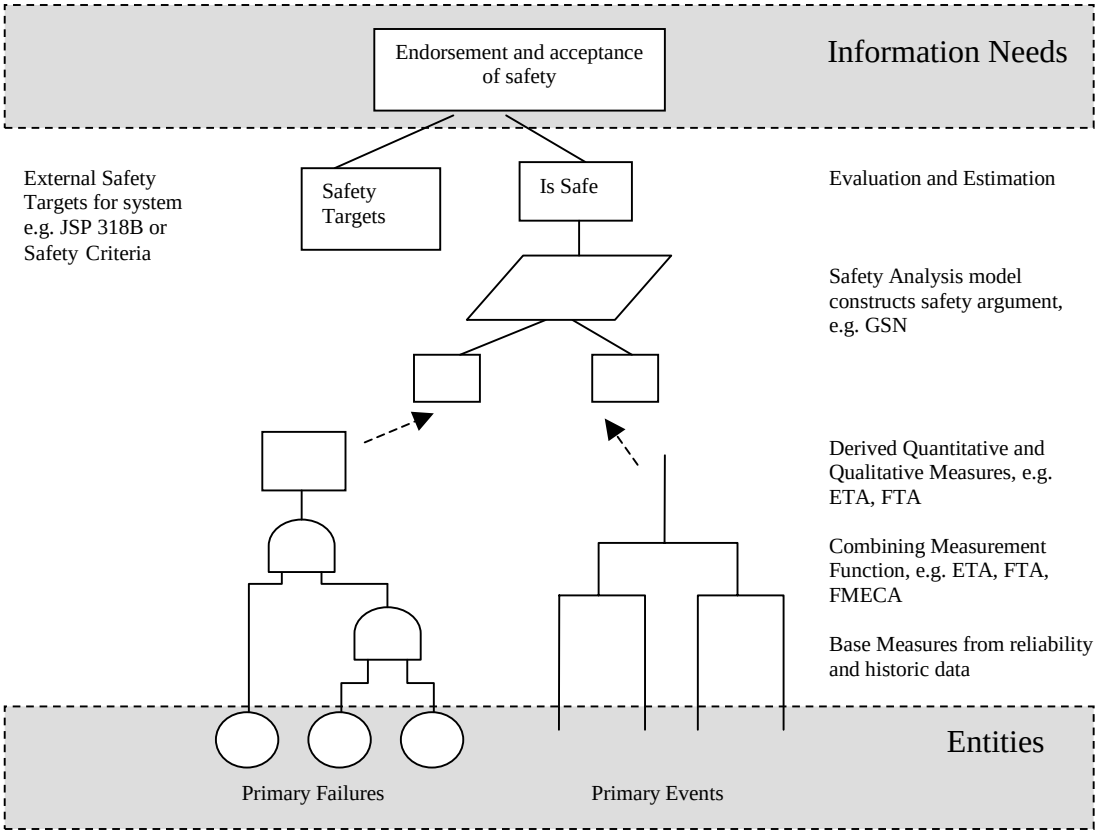

Figure 5-1 A measurement construct for a MOD safety case

Figure 5-1 A measurement construct for a MOD safety case is a variation on the PSM principle of measurement construction discussed by McGarry et al. This diagram simply illustrates how safety evidence base measurement entities map to safety information needs. In this particular variation the whole safety programme can be viewed as a measurement construction, where base measures (failures, events) are transformed by measurement functions (FTA, ETA) into derived measures and analysed Goal Structured Notation (GSN) to produce indicators for comparison against numerical thresholds (safety targets).

PSM is widely accepted industry measurement framework and many of the concepts are easy to implement, e.g. I-C-M tables and explaining measurement constructions.

5.3 Measurement of Organisations

For some organisations, safety issues dominate their business to such an extent that they must measure and demonstrate safety to a third party - usually a regulatory requirement, e.g. Nuclear. This has led to the development of organisational assessment schemes some of which factor in assessment of the design and quality of safety processes. Other factors such as competency and training assessment are also important – these also impact on the quality and effectiveness of safety processes.

All of these systems aim at assessing the overall organisations safety management system. Some employ a checklist approach, using safety audits and major accidents, Hurst et al [51] created a ten-point checklist scheme based on AIChe terminology and processes defined in the chemical plant industry. The work was extended in Harrison [49] and also Hurst and Radcliffe [50]. This latter methodology being based on a five-tiered sociotechnical pyramid model of causes of accidents.

Recent research by Wilpert and Rainer [52] identified thirteen differing methodologies of safety organisational assessment including that of Hurst and Radcliffe. Wilpert and Rainer proposed a rationalisation of all the techniques to produce a common approach based on seven generalised organisational factors. They recognise that there was considerable overlap in these factors and that many were not independent.

From this survey, four of the thirteen surveyed methods assessed parts of the organisation using incident and accident analysis techniques to identify weakness in the nuclear plant systems, management and procedures. So this leads to the question, whether the event analysis processes should be measured in order to determine the validity of the analysis results, i.e. can you trust a measure if you cannot trust the measuring device.

Many of the surveyed assessments also factor in competency and training of staff, yet training and competency were allocated to different organisation factors. Such assessments are combined and supported by schemes such as the IEE/BCS Competency Guidelines [54] that give ordinal grades to individuals performing specific safety tasks. Thus, one measure could be used to replace at least two measured attributes in generalised organisational factors.

Two of the goals highlighted by Wilpert and Rainer were the integration organisational factors into techniques that support probabilistic safety analysis and the prediction of organisational change on safety. Given that few, if any, real measures were identified this seems to be optimistic. The final report compiled by Baumont et al [53] fails to follow on and consolidate the attempts to standardise on a common approach and ends in a wish list of research topics related to the organisational assessment.

Another form of organisational measurement is Capability Maturity Model integrated (CMMI^{SM} [56]) for software and systems engineering companies. Recent work for the Australian Defence Materiel Organisation by the Software Verification Research Centre at the University of Queensland has extended CMMI model to include safety [57]. CMMI assesses system and software key processes areas of organisations the Australian safety extensions add two categories of safety management and safety engineering. The goals of these extensions, shown in table 5-1, are high level. +SAFE identifies safety products such as safety arguments, safety cases and hazard logs.

The +SAFE work is currently being enhanced to include security as well as safety. This work is being supported by Industry and government organisations (US and Australian) with the aim of incorporating the extensions into CMMI.

CMMI[SM] Categories	Safety Process Areas	Goals
Project Management	Safety Management	Develop Safety Plans
		Monitor Safety Incidents
		Manage Safety-related Suppliers
Engineering	Safety Engineering	Identify Hazards, Accidents and Sources of Hazards
		Analyse Hazards and Perform Risk Assessment
		Develop Safety Requirements
		Apply Safety Principles and Requirements
		Support Safety Acceptance

Table 5-2 +SAFE extensions to CMMI

A feature of CMMI is that measurement is an integral part of organisation and that it should identify areas of process improvement. This implies that those that implement CMMI extensions should measure safety and security processes and the their wide acceptance in conjunction with PSM could mean that they become the standard safety measurement framework.

A specific safety oriented organisational measurement scheme is that of the Conformity Assessment of Safety Related Systems (CASS) functional capability assessment [55] that assists in accredited certification to IEC 61508. CASS is administrated by non-profit company which is still developing assessment schemes in order to fully assess products for certification against IEC 61508. However the first step was to produce an assessment scheme for the certification of an organisations functional safety capability – this is basically the organisations' capability to produce a safe product or service.

The functional assessment methodology is very similar to that carried out for the ISO 9000 quality standard. That is, a third party group of CASS assessors, from an accredited certification body (for UK these bodies are managed by the United Kingdom Accredited Service), assesses functional safety procedures and facilities of a company against the IEC 61508 standard. Other similarities to the ISO 9000 model include a scope statement which identifies the bounds of the assessment and the assessed companies capability, e.g. SIL 3 software for petrochemical industry. The functional assessment requires the CASS assessors to audit and assess "targets of evaluation" against predetermined IEC 61508 evaluation criteria, typically:

TOE Ref	FSCA TOEs Target of Evaluation	Purpose of TOE	Referring IEC 61508 Clause & Tables	Comments
13	Procedures for maintaining information on hazards with respect to Safety-Related Systems	To define the procedures for maintaining accurate information on potential hazards and safety-related systems;	1/6.2.1 a)	Assessor should review the Hazard and Risk analysis report

Table 5-3 Extract from Functional Safety Capability Assessment [55]

The CASS system assesses the whole organisation including aspects of safety culture. It provides no grading, unlike CMMI, you either pass or fail, but does assess against a widely accepted international standard. The scheme is available internationally, however, similar competitive initiatives are also being offered, e.g. the TÜV Management of Functional Safety.

5.4 Measuring Safety Personnel

The Fenton and Pfleeger model of software measurement identifies that resources should be measured as well as products and processes. The most influential resources that affect safety processes are the safety specialists. The competency of safety specialists is an issue that requires assessment by some standards, e.g. IEC 61508 part 1 Annex B. This has resulted in the IEE/BCS Competency Guidelines [54] which is a staff assessment scheme for organisations involved in developing, manufacturing, operating and maintaining equipment built to standards such as IEC 61508.

The Competency guidelines provide a scheme for organisations to assess personnel against twelve different job roles (described as functions). These range from management roles at director level to detailed design levels such as software coding. For each job role, up to Fifteen separate job tasks are described and graded on a three level system of, supervised practitioner, practitioner and expert.

Some schemes assess the skill and experience of safety specialists against a specific engineering domain, e.g. the UK Institute of Railway Engineers licensing scheme. Others offer professional certification against generic safety, e.g. the Certified Safety Professionals sponsored by a number of US based organisations including the American Society of Safety Engineers.

5.5 Measuring Safety using Bayesian Belief Networks

Bayesian Belief Networks are described as directed acyclic graphs. The nodes in the graph represent probabilities associated with variables of interest (a set of events) and the arrows represent influences, causes or dependencies (this depends on your point of view, as there is some debate among researchers). The use of a graphical network allows intuitive reasoning on its validity, which is underpinned by the sound mathematical basis of Bayesian probability. Diverse evidence both qualitative (expert judgements) and quantitative can be mixed to give predictions. Perhaps more importantly, depending on the confidence of the model, it is possible to work backward through a network and establish the most important influences on a particular property and subsequently direct process improvement.

Extensive tool support with increases in computing power has made this technique practical allowing relatively large BBNs models. Consequently this is a popular research topic with a plethora of BBN models in many areas where risk or prediction is an important factor. BBNs are widely used to software to predict quality [58] and defects [59]. Fenton, Krause and Neil [60] predict that BBN technique:

"is the dawn of an exciting new era for software measurement."

BBNs have been also been used in safety related tasks, examples being:

- Assessment of nuclear safety Esprit DeVa project [61] and [62].
- Assessing dependability of safety critical systems the DATUM project [63].
- An attempt at using BBNs as a safety argument for a software safety case in the SHIP project [64].

The flexibility of BBNs in assessing processes is demonstrated by Woods [65] in a practical application where safety for software (ALARP) is measured by utilising cost benefit analysis techniques combined with BBN models. Woods created a general model for software that predicts the software failure rate but included diverse evidence from elements of software process, organisational pedigree, safety culture, personnel quality. The BBN modelling predicts the software probability of failure and considers the effect of specific hazards in a secondary BBN model. Cost benefit analysis is used to calculate the benefit of employing further safety risk reduction techniques in the software development, e.g. static code analysis.

It is clear that BBNs could provide quality, trust and effectiveness measures for a safety process. However, the diversity of the possible models and the amount of "expert" opinion/evidence make this a significant challenge, although one worthy of pursuing. In all the above BBN references validating the model has been a problem with authors referring to the models requiring many "iterations" with "trial and error" and better "evidence". Typical of the type of problems is those documented by Grup and Bosch [66] such as:

- the model is wrong (some models are very complex),
- more input evidence (the lack of quality quantitative evidence);
- evidence not matching reality (much of the evidence is based on opinion);
- and definition of variable terms.

This last problem is perhaps a significant concern as a node may be named "weather conditions" but these conditions could vary depending on the expert consulted, e.g. from Siberia or the Sahara. Perhaps a measure of how much trust we can place in these networks needs to be derived. Such a measure could be related to their complexity and context, however, verifying such a measure would be very difficult.

5.6 Summary - Other Methods and Relevant Work

GQ(I)M is a excellent mechanism for deriving information needs for a particular business goal. PSM is a software based measurement framework with extensive industry support which has been recently enhanced to include system elements including safety, although these latter elements are not well developed. PSM supports the measurement programs expected in the two most widely supported generic organisational assessment schemes, CMMI and ISO 9000. Recent extensions to CMMI have allowed the organisational assessment of safety management and engineering process but not its detailed processes.

Another noteworthy organisational assessment scheme is CASS. It assesses the safety process and management aspects of an organisation against IEC 61508. A significant number of other organisational safety assessment schemes exist, particularly in the Nuclear industry, but there is no dominant or universally accepted method.

Many of the organisation assessment schemes include factors such as training and experience. A noteworthy attempt is that of the IEE/BCS competency guidelines, which assesses up to fifteen different attributes for particular safety functions (jobs). Such a scheme could be factored into a measurement programme as a measure of resource quality.

BNNs are a flexible graphical mathematical technique that allows quantitative and qualitative data to be modelled. Creating a model to estimate safety process quality and effect on a system should be feasible but requires much effort in order to validate. Such a model would benefit from or depend on quantitative data taken from a safety process measurement programme.

6 Summary and Conclusions

This Review has investigated whether there are mechanisms to measure the quality, effectiveness and effort of safety processes. It has also investigated possible measurement frameworks for safety process improvement.

The definition of safety terms is diverse and the identification of useful measurable entities and attributes within safety processes is very difficult. Safety processes can be applied to many differing lifecycle standards and also include any activity that produces products for a Safety Management System. The scope of measuring safety processes is large and possibly more difficult than software process measurement. The safety process measurements surveyed are few, but those examined are diverse relating to organisations, investment and effort.

Effort measurements are needed for prediction and improvement comparisons. Basic effort estimation data for safety activities was, at best rudimentary, but PEL for safety shows promise in that it can overcome diverse nature of safety activities and their terminology. It may also allow useful comparisons between projects.

PSM, a software based measurement framework, could be adapted to include system safety elements. An additional advantage of PSM is that it supports CMMI and ISO 9000. Furthermore, safety extensions to CMMI are defined for the safety management and engineering process. Therefore, an adaptable measurement framework (PSM) and safety context organisational framework (CMMI) for process improvement exist for safety process measurement, albeit not fully developed or mature. The combination of PSM and CMMI appears to be the most promising safety process measurement framework.

Other useful measurement techniques that could be adapted are the IEE/BCS Competency Guidelines for assessment of human resources and Bayesian Belief Networks for estimating safety process quality and effectiveness. However, the development of a BBN is likely to depend on, or benefit from, quantitative data taken from a safety process measurement programme. Creating a BBN model to estimate safety process quality and effect on a system should be feasible but requires much effort in order to validate.

7 List of References

1. Defence Standard 00-56. Safety Management Requirements for Defence Systems. Issue 2, 13 December 1996.

2. Defence Standard 00-54. Requirements for Safety Related Electronic Hardware in Defence Equipment. Issue 1, 19 March 1999.

3. Defence Standard 00-55. Requirements for Safety Related Software in Defence Equipment. 1 August 1997.

4. Defence Standard 00-58. HAZOP Studies on Systems Containing Programmable Electronics. Interim Issue 1, 26 July 1996.

5. HSE, Reducing Risk, Protecting People, Health and Safety Executive, December 1999.

6. Joint Service Publication 454. "Procedures for Land Systems Equipment Safety Assurance". Issue 2, January 2000.

7. Fenton, N. E. & Pfleeger, S. L. Software Metrics, 2nd Edition, 1997, PWS.

8. Meulan, Meine van der. Definitions for Hardware and Software Safety Engineers, Springer, 2000, ISBN 1-85233-175-5

9. Lees, Frank, P. Loss Prevention in the Process Industries, Vols 1 and 2, Butterworth Heinemann, 1980, ISBN 0 7506 1529 X

10. Finkelstein, L.& Leaning, M. S. A Review of the Fundamental Concepts of Measurement, Measurement, vol. 2, pp. 25-34, 1984.

11. Stevens, S. S. Measurement, Psychophysics and Utility, Chapter 2 from Churchman C.W. and Ratoosh P. (Editors) Measurement: Definitions and Theories. 1958, John Wiley

12. McGarry, J.; Card, D.; Jones, C.; Layman, B.; Clark, E.; Dean, J.; Hall, F. Practical Software Measurement: Objective Information for Decision Makers. 2001, Addison-Wesley, ISBN 0-201-71516-3.

13. Park Robert E, Goethert, Wolfhart B.; Florac, William A. Goal Driven Software Measurement – A Guidebook (CMU/SEI-96-HB-002, ADA313946), 1996. Software Engineering Institute, Carnegie Mellon University.

14. Florac, William A.; Park Robert E.; Carleton Anita D. Practical Software Measurement: Measuring for Process Management and Improvement, 1997, Guidebook CMU/SEI-97-HB-003, Software Engineering Institute, Carnegie Mellon

University sponsored by US Department of Defence, and obtained through Department Technical Information Centre. (www.psmsc.com).

15. Card, David N.; Emam, Khaled El.; Scalzo, Betsy. Measurement of Object-Oriented Software Development Projects. 2001. Software Productivity Consortium, Sponsored by the Department of Defence. (www.psmsc.com).

16. Mil-Std-498, Software Development and Documentation, Department of Defence. December 1994.

17. ISO/IEC TR 15504:1998, Information Technology – Software Process Assessment.

18. Armitage, James W. & Kellner, Marc I. A Conceptual Schema for Process Definitions and Models, 53-165. Proceedings of the 3rd International Conference on the Software Process . Reston, Va., Oct. 10-11, 1994. IEEE Computer Society Press, 1994.

19. Caseley, P.R. Clark G.D, and Powell, A.L, White box measurement of through-life safety processes, MOD Safety Assurance Symposium, MOD Abbeywood, Bristol, 2001.

20. ISS, Introduction to System Safety, The University of York, Computer Science Department, Lecturers: McDermid, John; Kelly, Tim; Nicholson; Pumpfrey, David. 4-9 October1999

21. American Institute of Chemical Engineers (AIChE), Hazard Evaluation Procedures Second Edition with Worked Examples, 1992.

22. MIL-STD-882C. Military Standard, System Safety Programme Requirements. 19 January 1993.

23. Tribble, A.C. A Comprehensive Model of System Safety: A Tool for Determining Return on Investment. Proceedings of 18th International System Safety Conference. Fort Worth, TX: System Safety Society, 2000. 182-191.

24. Leveson, N. G. Safeware: System safety and computers. University of Washington, Addison-Wesley, 1995.

25. IEC 61508, Functional safety of electrical/electronic/programmable electronic safety-related systems, 1998.

26. ARP 4754. Certification Considerations for Highly-Integrated or Complex Aircraft Systems. 1996.

27. Murdoch, J., McDermid, J.A., Wilkinson, P. Failure Modes and Effects Analysis (FMEA) and Systematic Design. 19th International System Safety Conference, 10-14 September 2001, Huntsville.

28. Clark, Graham, D. and Powell Antony. L, Collaborative Software Process Data Collection: BAe and Roll-Royce's Experience of Sharing a Measurement Philosophy, Conference on Systems Engineering (INCOSE'99), Brighton, UK, 1999.

29. Clark, Graham, D. and Morris, Peter W.G, The development and application of a Process Engineering Language for project management data collection, ??? Note: have copy but not its publication details.

30. Powell, A. L, Right on Time: Measuring, Modelling and Managing Time-Constrained Software Development, University of York, Department of Computer Science, 17/08/2001.

31. QinetiQ/KI/SEB/CR020571/1.0, The effect of measurement for Safety Processes, available from the Defence Research Information Centre, MOD, 2001.

32. Jervis, Susan, and Collins, R. Terry, Measuring Safety's Return on Investment American Society of Safety Engineers, Professional Safety, September 2001.

33. Saaty, T. L., The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, McGraw-Hill,1980

34. Suokas, J. On the reliability and validity of safety analysis, 1985, Espoo, Technical Research Centre of Finland.

35. Rouhianinen, V. Importance of the quality management of safety analysis. Reliability Engineering and System Safety, vol 40, pp 5-16, Elsevier, 1993.

36. Harrison, Lee. Environmental Health, and Safety Auditing Handbook 2nd edition, McGraw-Hill, 1995, ISBN 0-07-026904-1

37. Rouvroye, J. L., van den Bliek, E. G., Comparing safety analysis techniques, Reliability Engineering and System Safety 75, 289-294, 2002

38. V. R. Basili and H. D. Rombach. The TAME Project; Towards Improvement Oriented Software Environment, IEEE Transactions on Software Engineering, vol. 14, pp. 758-773, 1988.

39. Bache, R. and Neil, M. Introducing metrics into industry: a perspective on GQM. In Software Quality Assurance Metrics, Edited by Fenton, Whitty and Iizuka, International Thompson Press, pp 59-68, 1995.

40. Wearing, A. Software Engineering, Ada and Metrics. (1992). Lecture Notes In Computer Science 603: 35-46

41. Park, Robert E., Goethert, Wolfhart B., Florac, William A. Goal-Driven Software Measurement —A Guidebook, August 1996. CMU/SEI-96-HB-002, Software Engineering Institute, Carnegie Mellon University.

42. Smith, David L. Using Measurement to Assure Operational Safety, Suitability, and effectiveness (OSS&E) Compliance for the C2 Product Line. 2001, 11th INCOSE Jul 2001 Melbourne.

43. ISO 9000:2000. Quality Management and Quality Assurance Standards.

44. Paulk, Mark C., Curtis, Bill, Chrissis, Mary Beth and Weber, Charles V. "Capability Maturity Model for Software, Version 1.1", Software Engineering Institute, CMU/SEI-93-TR-24, DTIC Number ADA263403, February 1993.

45. PSM Version 3.1a. Practical Software Measurement, A Foundation for Objective Project Management, Apr 1998, Office of the Undersecratary of Defence for Acquisition and Technology, Joint Logistics Commanders, Joint Group on Systems Engineering. (obtained from www.psmsc.com)

46. PSM Version 4.0b Draft. Practical Software and System Measurement, A Foundation for Objective Project Management, Oct 2000, Department of Defence and US Army,. (obtained from www.psmsc.com)

47. USAF, AFPD 63-12, Assurance of Operational Safety and Effectiveness, dated Mar 2000.

48. ISO/IEC 15939 Draft Standard, Software Measurement Processes, Jan 2001 (final co-ordination draft).

49. Harrison P.I.. Organisational, Management and Human Factors In Quantified Risk Assessment. Report 2, HSE Contract Research Report, No 34/1992.

50. Hurst, N.W & Radcliffe, K.B. Development of a Structured Audit Technique for Assessment of Safety Management Systems (STATAS), Instituted of Chemical Engineers Symposium Series No. 134, pp315-339, 1993.

51. Hurst, N.W., Bellamy, L.J. & Geyer, Y.A.W. … Organisational Management and Human Factors In Quantified Risk Assessment: a theoretical and empirical basis for modification of risk estimates, In M.H. Walter and R.F. Cox (editors), Proceeding of the Safety and Reliability Society, 1990, Altrincham, pp70-79, Elsievier Applied Science

52. Wilpert, B. Miller, R, Organisational Factors, Their definition and influence on nuclear safety (ORFA) – Report on needs and methods. Commission of the European Communities Forth Framework Programme on Nuclear Fission Safety, Contract No ERB FI4S-CT98_0051, 1999.

53. Baumont, G., Wahlström, B., Solá Ciemat, R., Williams, J., Frischknecht, A., Wilpert, B., Rollenhagen, C. Organisational Factors, Their definition and influence on nuclear safety – Final Report. Commission of the European Communities Forth Framework Programme on Nuclear Fission Safety, Contract No ERB FI4S-CT98_0051, 2000.

54. IEE/BCS. SAFETY, COMPETENCY AND COMMITMENT Competency Guidelines for Safety-Related System Practitioners, The Institution of Electrical Engineers, 1999.

55. CASS, The CASS Guide: Guide to Functional Safety Capability Assessment, Accredited certification to IEC 61508. 26 April 2000, Issue 2a. http://www.case.uk.net.

56. CMMI Development Team, Capability Maturity Model – Integrated System/Software Engineering (Version 1). Software Engineering Institute, Carnegie Mellon University. 2000.

57. Document ID: CA38809-364, +SAFE - A Safety Extension to CMMISM , Report to Australian Department of Defence, Dec 2001.

58. Neil, Martin and Fenton, Norman. Predicting Software Quality using Bayesian Belief Networks, Proceeding of 21st Annual Software Engineering Workshop NASA/Goddard Space Flight Centre, December 4-5, 1996

59. Fenton, Norman., Krause, Paul and Neil, Martin. A Probalistic Model for Software Defect Prediction, under revision for IEEE. Trans. Software Engineering, http://www.dcs.qmul.ac.uk/~norman/papers.html, 2001

60. Fenton, Norman., Krause, Paul and Neil, Martin. Software Measurement: Uncertainty and Causal Modelling, IEEE Software 10(4) 116-122, but available from http://www.dcs.qmul.ac.uk/~norman/papers.html, 2001

61. Fenton N.E., Littlewood B., Neil M., Strigini L., Wright D.R. (City University, London) and Courtois P.-J. (AVN, Brussels), Bayesian Belief Network Model for the Safety Assessment of Nuclear Computer-based Systems, DeVa ESPRIT Long Term Research Project No. 20072 - 2nd Year Report, pp. 485-512, Dec, 1997. Available from http://www.csr.city.ac.uk/people/lorenzo.strigini/ls.papers/DeVa_BBN_reports

62. Littlewood B., Strigini L., Wright D. (City University, London) and Courtois P.-J. (AVN, Brussels) Examination of Bayesian Belief Network for Safety Assessment of Nuclear Computer-based Systems, DeVa ESPRIT Long Term Research Project No. 20072 - 3rd Year Report, pp. 411-448, Dec. 1998.

63. Fenton NE, Littlewood B, Neil M, Strigini L, Sutcliffe A, Wright D, Assessing Dependability of Safety Critical Systems using Diverse Evidence, IEE Proceedings Software Engineering, vol. 145(1), pp. 35-39, 1999.

64. Delic K. A. , Mazzanti F. and Strigini L., Formalising a software safety case via belief networks, SHIP Project Technical Report T046, 1995.

65. Woods, Stewart. A Pragmatic Application of the ALARP Principle To Software, University of York MSc Dissertation, September 1999. www.cs.york.ac.uk/MSc/SCSE/local/projectabstracts.html.

66. Gurp Jilles van., Bosch, Jan, Using Bayesian Belief Networks in Assessing Sofware Architectures, ICT Architecture in the BeNeLux 1999 (ICT-Architecture'99) November 18-19, 1999, Amsterdam.

Initial Distribution

Mr. Ed Swindle, MoD customer

Mr. Mike Simpson, Dstl Programme Director

Mr: Gareth Rowlands, DPA Ship Safety Management Office

Sqn Ldr Mike Musslewhite, DPA ADRP

Mr. John Erbetta, MoD customer

Mr. Mike A Simpson

Mr. Tom McCutcheon

Dr. Tim Thorp

Mr. John MacRae (QinetiQ)

Ms. Cherly Jones, PSM Project Manager, US Army TACOM-ARDEC

Mr. Joe Jarzombek, Software Intensive Systems, AT&L, OSD

Mr Matt Ashford, SIS, AT&L, OSD

Report Documentation Form

A copy of this form is to be completed by the principal author for all Dstl reports. When complete, it is to be bound as the last numbered pages of the finished report.

1. Originators Report Number incl. Version No DSTL/CP06715 V1		
2. Report Protective Markings and any other markings e.g. Caveats, Descriptors, Privacy markings None		
3. Title of Report Safety Process Measurement – A Review		
4. Title Protective Markings incl. any Caveats None - unlimited		
5. Authors Report		
6. Originator's Name and Address Paul Caseley N133 DSTL Malvern St. Andrews Road Malvern WR14 3PS UK	7. MOD Sponsor Name and Address Edward Swindles Communications, Information and Signal Processing (CISP) A3 Room 25 DSTL FortHalstead	
8. MOD Contract number and period covered 2002/3		
9. Applied Research Package No. N/A	10. Corporate Research Package No. TG10 ND/10/05/03/013	11. Other Report Nos.
12. Date of Issue May 2003	13. Pagination 38	14. No. of References
15. Abstract (A brief (approximately 150 words) factual summary of the report) This reports reviews current and past research into safety process measurement. The report also discusses the basic measurement principles, problems associated with measuring safety processes, measurement frameworks and current industry practices for measuring system development processes.		
16. Abstract Protective Marking including any Caveats Unlimited		
17. Keywords/Descriptors (Authors may provide terms or short phrases which identify concisely the technical concepts, platforms, systems etc. covered in the report. Safety, measurement, processes, metrics		

18. Report Announcement (refers to title/abstract being included in accessions lists e.g. Defence Reports Abstracts)

☒ Announcement of this report is UNLIMITED

If there are limitations on the announcement of this report please indicate below the groups to whom it can be announced (more than one if required)

☐ Can be announced to MOD and its Agencies

☐ Can be announced to UK Defence Contractors

☐ Can be announced to Overseas Defence Departments

☐ Other (please specify)

19. Report Availability

☒ UNLIMITED distribution

☐ No Release without approval of the Release Authority

If the above do not apply, please indicate below the groups to whom the report may be released upon request without further Need-To-Know checks.

☐ Can be released to MOD and its Agencies

☐ Can be released to other UK Government Departments

☐ Can be released to UK Defence Contractors

☐ Can be released to Overseas Defence Departments

☐ Other (please specify)

20. Downgrading Instructions (check as appropriate)

☐ This report may be automatically downgraded to after years

☐ This report may be reviewed years after publication

21. Authorisation (Complete as applicable)

	Name	Signature	
Project Manager	John Evans QinetiQ		Date
Technical Reviewer	Robertn Anderton		Date
Customer			Date

When complete the form is to be bound into the report to which it refers and is to form the last numbered pages of the report. Dstl Knowledge Services, Glasgow will enter an abstract and other details onto the relevant report management systems.

This page is intentionally blank