THE STATE OF PRACTICE IN SYSTEM SAFETY RESEARCH EVALUATION

A.J. Rae, M. Nicholson, R.D. Alexander

University of York, UK, ajrae@cs.york.ac.uk, mark@cs.york.ac.uk, robert.alexander@cs.york.ac.uk

Keywords: Research Practice, Evaluation, Methodology.

Abstract

This paper reports an investigation into the use of evaluation as part of system safety research. Using a simple classification scheme based on the knowledge and evaluation components of research papers, we classify two years of papers at the IET System Safety conference. Our analysis indicates a significant mismatch between a small collection of observational research papers with strong evaluation, and a large body of papers providing guidance which have not been evaluated. Of particular concern is that the majority of these papers do not provide sufficient information to support future evaluation. In response to these findings we suggest a minimum set of properties which guidance research products must have in order to allow the research to be evaluated.

1. Introduction

1.1 The Importance of Evaluation

Pure science attempts to describe, classify and explain the world as it is. In contrast, engineering and applied science (categories which include system safety) primarily seek to generate ideas which add to or change the world. Such ideas may be designs of concrete objects or designs of ways of doing things. Either way, they should be thought of as deliverable research artefacts, with an implicit requirement that they be fit for purpose.

We propose that system safety research products are fit for purpose if they engender positive change in industrial practice for the development, deployment, operation and disposal of safety critical systems. As with any product, if we are to be confident that it is fit-for-purpose before it is applied we must do some predictive analysis. Even once it is used, we need to assess whether its effect on safety is actually positive; given the difficulty of measuring safety, this is by no means guaranteed. We therefore need research evaluation.

It's important to note that measuring "impact" alone is not sufficient – a technique could be widely used, but have flaws that are not recognised. This is perhaps particularly likely in system safety, because safety is so difficult to measure.

Similarly, whilst peer review is an important form of evaluation, it is not a measure of efficacy. As we discuss in Section 4.5, peer review can support other forms of evaluation but cannot replace them.

1.2 Evaluation in Systems Safety Research

The authors, as practitioners, educators and researchers, have formed an impression that whilst safety engineering research generates a large number of plausible and potentially useful ideas, very few of these ideas are tested to determine their efficacy. Some techniques have clearly achieved success insofar as they are used in practice, but widespread adoption of a technique is not evidence that the technique is actually cost-effective in improving safety.

There is an important distinction here, which parallels the practice of safety engineering itself: just as a system may be safe but lack evidence of safety, so a technique may improve safety, but lack evidence of its efficacy. Evidence and efficacy are distinct but not independent concepts – lack of evidence always justifies scepticism about claims.

We think that the lack of knowledge about the relative efficacy of techniques is a potential barrier to advancement of the state of the art and the state of practice.

This paper documents our investigation of the first part of this concern, by investing how much evaluation of system safety methods and techniques is published in current system safety research.

2. Methodology

There are numerous frameworks for categorising knowledge (e.g. Holsapple and Singh [5]). In Alexander et al [2] we provide a more detailed discussion of knowledge types relevant to system safety and suggest research techniques applicable to each.

For the purposes of this paper, we are concerned with two specific types of knowledge. The first we term "observation". Observation characterises the past or current state of some portion of the real world. The second type of knowledge we term "guidance" which describes how to perform a particular action or set of actions.

We observe that this categorisation is not a complete description of all research, as there are valuable research contributions which do not provide observation or guidance.

It is also not exclusive, as a single research publication can provide both observation and guidance.

We used this framework to categorise the papers published in the proceedings of the 2008 and 2009 IET System Safety Conference. IET System Safety was selected because it is representative of conferences focussing on system safety.

For each paper, we recorded:

- Whether the paper explicitly created or reported observation knowledge
- Whether the paper explicitly created or reported guidance knowledge
- Whether the paper explicitly evaluated the knowledge that was created or reported
- Whether the paper explicitly evaluated knowledge from a different paper or papers

Our investigation made no judgement about the quality of the contribution of each paper, but relied on the explicit statements made within the paper about the nature of the reported research. For example, if any method was used to evaluate knowledge, we recorded that the paper contained evaluation – we did not make a judgement about the appropriateness of the method or the way it was applied.

We used two recorders (Rae and Nicholson). Most papers were classified by one person, but cross-recorder reliability was checked by using both reviewers on 25% of the papers. For this sample, the reviewers achieved high agreement as to the categories applied to each paper.

Prior to the study, we proposed three hypotheses for investigation:

1. That a significant majority of research papers which provided guidance would not also include evaluation of the knowledge presented.

2. That there would be an extremely small sample of research papers which sought to evaluate guidance presented in earlier research.

3. That in designing new guidance, most research relies on unpublished sources for observations of the real world. Actual published observations tend to be either exemplar case studies or accident analysis; they are rarely research about typical practice.

If our hypotheses are correct, then they suggest that future research which focuses on observation of practice and evaluation of existing guidance should be a priority.

Whilst we cannot claim detailed knowledge of how practitioners select a suite of practice, researchers cannot realistically expect practitioners to adopt new techniques without clear evidence of the appropriateness and effectiveness of those techniques. The idea that theoretical plausibility is a reasonable substitute for proven efficacy has been discredited in other fields [7]

As well as testing the hypotheses explicitly, we also searched informally for explanatory clues or trends as to the state of practice in research validation.

3. Results

Here, we summarise the results of our study; the full data (the list of papers coded for contribution types and evaluation content) is available from the authors on request.

3.1 Overview of Sample

Thirty-eight papers were reviewed from IET System Safety 2008, and forty-three papers from IET System Safety 2009.

Of this set:

- 11 papers contained no guidance, observation or evaluation (and were therefore out of scope for this study)
- 48 papers contained guidance and no observation
- 6 papers contained just case study observation
- 5 papers contained just other forms of observation
- 10 papers combined guidance and observation
- 1 paper contained just evaluation

This distribution is shown in Figure 1.

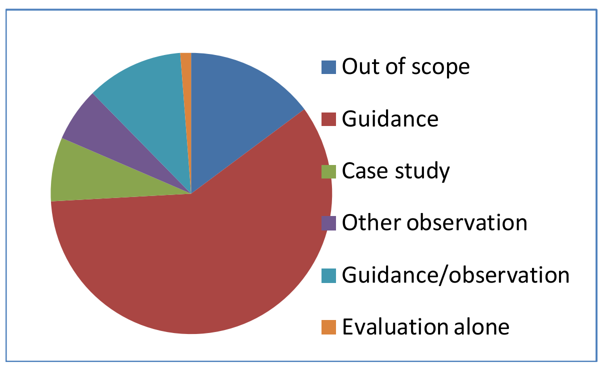

Figure 1: Distribution of Paper Classifications

The eleven papers not containing knowledge relevant to the study were a mix of literature surveys, issue discussion papers and progress reports on development of particular standards. One of the papers contained the design details of a specific artefact.

Of the relevant papers, eleven contained evaluation:

- 4 guidance papers which evaluated the guidance
- 1 case study paper, which evaluated the conclusions drawn from the case study
- 4 guidance and case study papers, which explicitly used the case study as a form of evaluation

- 1 paper which evaluated work previously published by others
- 1 paper which collected data, evaluated the conclusions drawn from that data, and used those conclusions to evaluate the effect of a regulation

3.2 Evaluation of Hypothesis 1

Hypothesis 1 was that a significant majority of the papers which contained guidance would not also evaluate that guidance. Of the fifty-eight papers classified as containing guidance, only eight papers provided any evaluation of the guidance provided, confirming the hypothesis.

We have not attempted to judge whether the evaluation method used is capable of providing useful or trustworthy information about the guidance; for the current work, we have assumed that all evaluation is equally good. In the longer term, we would like to study the range of evaluation techniques applied, and the rigour with which they are performed. We have made a start in Alexander et al [2], which discusses evaluation techniques appropriate to each of the knowledge types.

3.3 Evaluation of Hypothesis 2

It is not necessarily a problem if a paper presents un-validated guidance, but the research life-cycle is not complete until the guidance is evaluated. If the guidance is presented in one paper, and subsequent papers present evaluation of the guidance (by either the original authors or by others), this would be an indication of a healthy research community.

In the sample reviewed, **only one** paper evaluated previously reported guidance.

Of further concern, most of the papers in our sample actually lack the properties needed to allow future research to evaluate them. These properties are discussed in Section 4.4 below; when a paper does not have them, it is extremely difficult to evaluate it well.

3.4 Evaluation of Hypothesis 3

The third hypothesis was that in designing new guidance, most research relies on unpublished sources for observational knowledge. The data partially confirms this hypothesis, by showing a preponderance of guidance papers compared to observational knowledge papers.

A total of 58 papers contained guidance, whereas 21 papers contained observational knowledge. 12 of the observational papers were case studies, whereas 9 contained other forms of real-world description.

The data shows that whilst guidance and case studies appeared together in 6 papers, in only one of these cases was evaluation explicitly conducted.

Where other forms of observation appeared with guidance (4 papers), the observation was used to generate the guidance.

Classification	Total papers	# Evaluated
Guidance	58	5
Case Study	12	1
Other Observational	9	4
Guidance + Case Study	6	1
Guidance + Other Observational	4	0

Table 1: Evaluation of Papers

Table 1 shows the number of papers of each type which contain evaluation. This data shows that there is a much stronger focus on evaluation for papers containing observation other than case studies.

Direct comparisons cannot be drawn because the nature of evaluation is different for guidance, case studies, and more quantitative observation.

3.5 Other Observations

The remarks in this section are not shown directly in the data set, but are observations made by the authors whilst conducting the study.

Based on the sample we reviewed, it appears guidance papers follow one of three patterns.

The first pattern, which we could call the “brand new method approach”, is:

1. Describe a problem
2. Describe a method
3. Illustrate method with case study OR
4. Perform a combined illustration/evaluation by application of the method to a real world example

The second pattern, which we could call a “case study approach”, is:

1. Describe a particular real-world scenario
2. Describe the method followed in that scenario

The third pattern, which we will refer to as a “standard driven approach”, is:

1. Describe the requirements of a standard or regulation
2. Describe a gap in guidance available to follow that standard (which may be due to some new situation or technology)
3. Describe a method which fills the gap and is compliant with the standard
4. Illustrate the method with a case study

The implicit claim with all of these patterns is that the guidance provided has desirable outcomes which will be obtained by others if they follow the guidance. The claim is seldom stated explicitly, which makes it difficult for others to evaluate the research: it is possible to evaluate against a specific desirable outcome, but not to evaluate against all conceivable desirable outcomes. Either a paper states specific claims ("my guidance achieves X better than guidance Y, within application domain scope Z"), or anyone who wants to apply or evaluate it will have to infer such claims themselves.

It can be reasonably assumed that anyone publishing guidance or case studies intends to inform the research or practice of others, and non-explicit claims reduce the ability to inform.

4. Discussion

4.1 Internal Validity of the Study Methodology

Our classification scheme involved the use of human reviewers, which made it possible for reviewer error to distort the measurements. In particular, a study of the quality or effectiveness of evaluation would have been highly dependent on the competence and philosophy of individual reviewers.

Our study was designed to minimise any such effect; we relied on explicit statements made within the papers about whether any evaluation had been performed. This will have reduced the subjectivity of the measurements.

The inter-reviewer reliability of the classification scheme (see Section 2) increases confidence in the success of this approach.

4.2 External Validity of the Study Methodology

The IET conference was selected for this review because we believe it to be representative of publication in the field. This is an untested assumption, and to confirm that the conclusions can be generalised we will need to replicate the results for other publication venues.

We can, of course, only measure public domain research, and so the approach taken in this study cannot draw conclusions regarding evaluation which was conducted but not reported.

4.3 Requirements for Evaluation

For each type of knowledge, there is a benchmark that distinguishes those things that are reasonable to believe in. Meeting the benchmark does not compel belief, but research which fails to meet the benchmark makes no meaningful contribution.

The process of measuring research is evaluation. We are specifically concerned here about evaluating the truth of research outcomes. We are ignoring other dimensions, such as the "impact" proposed for inclusion in the Research Excellence Framework (REF) [1], although as noted earlier we believe that better-evaluated research will have greater impact.

It is not necessary that any single paper or even any single project includes evaluation, although the research life-cycle is incomplete until evaluation has occurred. It *is* necessary that any knowledge provided by the research has the properties that are needed for it to *be* evaluated.

Observations are evaluated by determining how well they correctly report the state of the world, and whether any conclusions drawn are necessary and sufficient explanations of the evidence. Thus, evaluation of such research is concerned with things such as sample sizes and design of suitable controls.

Guidance is evaluated by showing that it can be followed, and that following the guidance brings about some desirable result.

4.4 Guidance on Well-Formed Guidance Knowledge

In order to verify that evaluation can be performed on a guidance paper, we provide the following four-part test:

1. The scope within which the guidance is intended to be applied must be clear.

2. The guidance must be precisely and explicitly stated. Ideally, they will be stated in multiple consistent ways (e.g. as a workflow and as a worked example).

   This part of the test may be achieved without publishing excessive detail – citation of an external source such as a published technical report, thesis or handbook meets the intent of this requirement.
   In the case of changes to a previously published method, they can be expressed simply as a delta to a description elsewhere.

3. Where the guidance is generic about any element, at least one instantiation of the element must be provided. For example, if a proposed life-cycle includes "hazard identification", the authors must name at least one hazard identification technique that is suitable.

4. There must be claims about the benefits of following the guidance (in situations that are within the scope). These claims should be amenable to evaluation.

Whilst this test is sufficient to allow evaluation of the research, there is an additional "strength of claim" property necessary for knowledge transfer from research to practice.

5. The testable claims must be such that truth of the claims compels a change in practice.

The four elements of the test for well-formed knowledge may be considered as the "verification" component of evaluation. The strength property (test 5) is the "validation". Research which makes true but weak claims is not useful.

For example, there are a plethora of analysis techniques where the implicit claim is that they can find some errors in system models. Even if this claim is fully substantiated, it gives no reason to prefer one technique over any other technique also capable of finding some errors. Only a claim that a technique is *more capable* than another technique, within some defined scope, compels a change in industrial practice. “More capable”, here, can include being a cheaper way to produce the same quality of result.

4.5 Peer Review as a Form of Evaluation

Every research paper published in a major conference or journal undergoes at least some evaluation in the form of peer review before the paper is accepted. This is good, but evaluation can take many forms, and each type of evaluation is suited to measuring different research properties. An analogy can be drawn with various types of safety evidence for software:

- Audits can provide good evidence that processes were followed, but does not investigate product properties
- Dynamic testing can provide good evidence that functional requirements are met, but does not show that certain undesirable behaviours can’t happen given the wrong circumstances
- Static analysis can show conformance with specification, but can’t show conformance with designer intent

In the same way, each type of evaluation can only measure certain properties. In the case of standard peer review, it measures attributes of the reporting of research. As such, peer review has been shown to have strong effects in improving the quality of research reporting [3]. Peer review is also highly reliable in identifying serious methodological flaws [8]. It can, of course, check well-formedness properties such as those we gave in the previous section.

Peer review has some capability to predict research impact [8], but this is very reviewer-dependent, and there may be more focused methods of evaluation that can produce better predictions [4]

Standard peer review cannot measure the properties of a research artefact itself, as opposed to the research reporting. Peer review specifically does not attempt to replicate research or to test the veracity of its findings. Other forms of evaluation such as observational studies or deliberate tests and experiments are necessary to show that an applied science research artefact has specific properties.

5. Attacking the Problem

5.1 Further Investigation into the State of Practice

The data generated by this study is limited in scope, and the conclusions should not be generalised at this stage. If we want to propose changes to behaviour, we will need better knowledge of the situation.

We have an ongoing project to test our hypotheses against system safety research published in other venues. We are particularly interested in whether the prevalence and quality of evaluation varies with the publication venue.

Alongside this expansion of data, there is a need for matched control data. There are good reasons, as discussed by Alexander [2], why system safety research evaluation is difficult. Insight may be gained from looking at evaluation methods in systems engineering and software engineering research, because those fields face similar difficulties.

5.2 Investigation of Research Consumption

At present, little is known about how organisations make decisions about changes in system safety practice. Experience within the University of York, as documented for example in Wiseall [9], indicates that successful technology transfer of safety techniques may be driven by factors such as the commercial-readiness of tools and convincing safety engineers of the “right” way to practice safety, rather than by having strong evidence of the efficacy of the proposed techniques. There are also contradictory anecdotes, however, which suggest that lack of good evidence of benefit may prevent adoption of credible research.

If we are to improve the evaluation of system safety research, we will have to convince researchers that empirically-supported claims are necessary, and encourage industry to both demand and fund research which includes strong evaluation. This, in turn, will require knowledge about how companies make decisions regarding safety practice.

A likely driver of changed practice is response to perceived risk and to counter-evidence of practice efficacy. Accidents, incidents and safety-related failures, whilst undesirable in their own right can have positive effects on practice. Encouraging positive changes which pre-empt accidents, rather than respond to them is a desirable goal for this type of research.

5.3 Demonstrating Feasibility of System Safety Research Evaluation

Our record of the current state of research evaluation is unlikely to surprise most practitioners and researchers in the field. The main contribution of this paper is that we have identified specific characteristics of the problem which can be remedied.

We would like to be able to claim that shifting the focus of system safety research towards stronger evaluation *is* practical. Whilst we believe such a claim would be true, there is insufficient evidence to support it. Until such evidence exists, it is easy for research paper authors to say “guidance evaluation is a nice idea, but it’s not practical in our field”.

An important next step is to collect a set of compelling exemplar papers, each of which evaluates some aspect of system safety guidance. This will help to demonstrate that evaluation of such guidance research of this nature can be both practical and effective. In this endeavour, we will explore the adaptation of existing techniques for assessing research maturity, such as Technology Readiness Levels (TRLs) [6].

The authors also plan to carry out some small-scale empirical investigations. These projects will act as empirical evaluation of our own ideas about empirical evaluation.

5.4 Design of Education and Research Consumer Interventions

The authors are involved in teaching Masters-level degree courses in safety and software engineering, to both new graduates and experienced practitioners. Part of the education process includes student projects, which usually cover a complete research or product-development life-cycle. We may be able to improve our degree programmes by generating better standards for evaluation, and training students on how to meet those standards.

We plan to design and test different ways of teaching students about the importance of evaluation, and about selecting appropriate mechanisms for evaluation. This in turn we hope will have a direct effect on both the quality of future research, and on the students' ability to judge the research of others.

5.5 Implications for Research Sponsorship and Review

To the extent that the findings of our study reflect negatively on the current state of research practice, this should not be seen as a criticism of any of the specific research outcomes. The level of research evaluation is an emergent property of the research community, and not a flaw in any individual unit of research.

There are, however, ways in which we believe that research sponsorship and peer review organisations can support growth in the evaluation of research. In particular, we encourage the adoption of the kind of tests we describe in Section 4.4 as reviewer guidelines. We do not believe that this will have unwanted side effects.

Whilst research sponsors could favour projects which contain a significant evaluation component, the existence of evaluation is not a test which can be applied at the level of individual research papers – a paper doesn't need strong evaluation in order to be valuable. The important thing is that *eventually* all interesting work is evaluated somewhere.

Hence, we do not believe that evaluation of claims should be an *absolute* review criterion, although it may be appropriate for conferences to *encourage* papers that contain evaluation.

6. Conclusion

This research is unusual in that the subjects are, albeit indirectly, the members of the research community themselves, including the authors of this paper. We find cause for optimism in the observations we have made, as there appears to be a clear path towards improving the status of system safety research as both an aid to industry and an academic discipline.

The authors welcome feedback on the ideas presented here, and in particular would appreciate hearing the experiences of those who have been frustrated or unsuccessful in attempts to conduct or publish research evaluation in this field.

References

[1] Excellence Framework, "The Research Excellence Framework : A brief guide to the proposals What is the Research Excellence Framework ?," 2009.

[2] R. Alexander, A.J. Rae, and M. Nicholson, "Matching Goals and Methods in System Safety Engineering," *IET System Safety*, 2010.

[3] S.N. Goodman, J. Berlin, S.W. Fletcher, and R.H. Fletcher, "Manuscript Quality before and after Peer Review and Editing at Annals of Internal Medicine," *Annals of Internal Medicine*, vol. 121, Jul. 1994, pp. 11-21.

[4] N. Haslam, L. Ban, L. Kaufmann, S. Loughnan, K. Peters, J. Whelan, and S. Wilson, "What makes an article influential? Predicting impact in social and personality psychology," *Scientometrics*, vol. 76, Jul. 2008, pp. 169-185.

[5] C.W. Holsapple and M. Singh, "The knowledge chain model: activities for competitiveness," *Expert Systems with Applications*, vol. 20, Jan. 2001, pp. 77-98.

[6] J.C. Mankins, "Technology Readiness Levels," Apr. 1995.

[7] A. Shapiro and E. Shapiro, *The Powerful Placebo*, Baltimore: Johns Hopkins University Press, 1997.

[8] J.F. Sollenberger, "Editorial Peer Review: Its Strengths and Weaknesses.," *Journal of the Medical Library Association*, vol. 90, Jan. 2002, pp. 115-115.

[9] S.S. Wiseall, J.C. Kelly, and T.P. Kelly, "Transferring Design Research In To Rolls Royce," *International Conference on Engineering Design*, 2001.

Classification of Papers

General note: The columns represent classifications, not scores. This table should not be used to compare individual papers.

DOI: Digital Object Identifier for each paper. The DOI uniquely identifies each paper in the data set, and can be used to find the paper via www.doi.org

Guidance: Does the paper provide guidance? (1 = yes, 0 = no). Guidance does not have to be a complete method, but may refine a method provided elsewhere, or give instructions on how to apply a method provided elsewhere. A paper is considered to provide guidance if the abstract indicates so, or if one or more numbered sections of the paper primarily consist of guidance.

Case Study: Does the paper include one or more case studies? (1 = yes, 0 = no). A case study is a new description of a real-world system, project or event. Extracting details directly from a case study published elsewhere is not recorded as a case study. Examples that are labelled as hypothetical are not recorded as case studies.

Other Observ: Does the paper include observational knowledge not in case study form? (1 = yes, 0 = no). This category includes both experiments and systematic collation of data from multiple sources.

Internal Evaluation Does the paper use any method to measure the validity or strength of conclusions made within the paper? (1 = yes, 0 = no). If the paper reports or records an activity, and states that this activity constitutes evaluation, this is recorded as self evaluation. Also, if the paper makes explicit statements about validity or strength arising from an activity that appears to be evaluation, this is recorded as self evaluation.

External Evaluation: Does the paper use any method to measure the validity or strength of conclusions made within other papers? (1 = yes, 0 = no).

Notes: This field records extra information which may be helpful in interpreting the classifications. If a paper is not classified in any other column, the nature of the paper is recorded in this field.

DOI	*Guidance*	*Case Study*	*Other Observ.*	*Internal Evaluation*	*External Evaluation*	*Notes*
10.1049/cp.2009.1532	0	0	0	0	1	
10.1049/cp.2009.1536	1	0	0	0	0	
10.1049/cp.2009.1533	1	0	0	0	0	References external case studies
10.1049/cp.2009.1537	0	0	0	0	0	Issue discussion paper
10.1049/cp.2009.1534	1	0	0	0	0	
10.1049/cp.2009.1538	1	0	1	0	0	Data analysis leading to recommendations
10.1049/cp.2009.1535	1	0	0	0	0	
10.1049/cp.2009.1539	0	0	0	0	0	Issue discussion paper
10.1049/cp.2009.1540	1	0	1	0	0	Compiled observations leading to advice
10.1049/cp.2009.1543	0	0	0	0	0	Competency scheme / future work brief
10.1049/cp.2009.1541	1	0	0	1	0	Validation by industrial case study
10.1049/cp.2009.1544	0	0	0	0	0	Literature survey
10.1049/cp.2009.1542	1	0	0	0	0	
10.1049/cp.2009.1545	1	0	0	0	0	
10.1049/cp.2009.1546	0	0	0	0	0	Issue discussion and future work
10.1049/cp.2009.1549	1	0	0	0	0	
10.1049/cp.2009.1552	1	0	0	0	0	
10.1049/cp.2009.1547	1	0	0	0	0	
10.1049/cp.2009.1550	0	0	1	1	0	Data analysis
10.1049/cp.2009.1553	0	1	0	0	0	Project case study
10.1049/cp.2009.1548	0	1	0	0	0	Incident case study
10.1049/cp.2009.1551	1	0	0	0	0	
10.1049/cp.2009.1554	1	0	0	1	0	Evaluation by case study
10.1049/cp.2009.1555	1	0	0	1	0	Evaluation by experience with use
10.1049/cp.2009.1558	0	0	0	0	0	Literature survey
10.1049/cp.2009.1561	1	0	1	0	0	Data analysis leading to recommendations
10.1049/cp.2009.1556	1	0	0	0	0	
10.1049/cp.2009.1559	1	0	0	0	0	
10.1049/cp.2009.1562	1	0	0	0	0	

10.1049/cp.2009.1557	1	0	0	0	0	
10.1049/cp.2009.1560	1	0	0	0	0	
10.1049/cp.2009.1563	1	0	0	0	0	
10.1049/cp.2009.1564	1	0	0	0	0	
10.1049/cp.2009.1566	1	0	0	0	0	
10.1049/cp.2009.1568	0	0	0	0	0	Discussion of standard - effectively literature survey
10.1049/cp.2009.1565	1	1	0	0	0	Discussion of application of method
10.1049/cp.2009.1567	1	0	0	0	0	
10.1049/cp.2009.1569	1	0	0	0	0	
10.1049/cp:20080696	1	0	0	0	0	
10.1049/cp:20080697	1	1	0	0	0	Illustrated on real case study
10.1049/cp:20080698	1	0	0	0	0	
10.1049/cp:20080699	1	0	0	0	0	Literature review with recommendations
10.1049/cp:20080700	1	0	0	1	0	Framework validated by using it to analyse protocols and find bugs
10.1049/cp:20080701	1	0	0	0	0	
10.1049/cp:20080702	1	0	0	0	0	
10.1049/cp:20080703	1	0	0	0	0	
10.1049/cp:20080705	1	0	0	0	0	Reflects on need for evaluation
10.1049/cp:20080706	1	0	0	0	0	
10.1049/cp:20080707	1	1	0	0	0	Lessons learned from lit review + direct experience case study
10.1049/cp:20080708	0	0	1	1	0	Experimental study + consideration of limitations
10.1049/cp:20080709	0	1	0	0	0	Experience report including measure of effort expended
10.1049/cp:20080710	1	0	0	0	0	Notes difficulty of evaluating
10.1049/cp:20080711	1	0	0	0	0	
10.1049/cp:20080712	0	0	1	1	0	Qualitative data collection and analysis
10.1049/cp:20080714	0	1	1	0	0	Real life worked example - nb risk is "evaluated" but guidance is not
10.1049/cp:20080715	1	0	0	0	0	
10.1049/cp:20080716	1	0	0	0	0	
10.1049/cp:20080717	1	0	0	0	0	
10.1049/cp:20080718	1	0	0	0	0	
10.1049/cp:20080719	0	0	0	0	0	Progress report on standard preparation
10.1049/cp:20080720	0	1	0	0	0	Accident/Incident case studies
10.1049/cp:20080721	1	0	0	0	0	Includes comment that scheme accepted by regulator - not classified as evaluation
10.1049/cp:20080722	0	0	0	0	0	Issue discussion paper
10.1049/cp:20080734	1	1	0	0	0	Case study and lessons learnt
10.1049/cp:20080735	1	0	0	0	0	
10.1049/cp:20080736	1	0	0	0	0	
10.1049/cp:20080737	1	0	0	0	0	
10.1049/cp:20080738	0	0	0	0	0	Ongoing research and future plans
10.1049/cp:20080739	1	0	0	0	0	
10.1049/cp:20080740	1	1	0	1	0	Statistical evaluation of survey validity, and case study evaluation of application
10.1049/cp:20080741	1	0	0	0	0	
10.1049/cp:20080742	0	0	1	1	1	Data analysis used to evaluate a new regulation
10.1049/cp:20080724	1	0	0	0	0	
10.1049/cp:20080725	0	1	0	0	0	Accident/incident case study
10.1049/cp:20080726	1	1	0	0	0	Process applied to case study
10.1049/cp:20080727	1	0	0	0	0	
10.1049/cp:20080728	1	0	0	0	0	Issue discussion and guidance
10.1049/cp:20080729	0	0	0	0	0	This is an inter-standard mapping. Any guidance is implicit, so not classified as such.
10.1049/cp:20080730	1	0	0	0	0	
10.1049/cp:20080731	0	0	0	0	0	Algorithm design with limited detail
10.1049/cp:20080732	1	0	1	0	0	Guidance based on descriptive observation