Assessing safety culture: recent developments and lessons learnt

Michael Wright, Director; Rebecca Canham and Shona Watson, Senior Consultants; Greenstreet Berman Ltd, 10 Fitzroy Square, Fitzrovia, London, W1T 5HP. www.greenstreet.co.uk

It is widely accepted that an effective safety culture is an essential enabler of effective safety performance. Over the past decade many safety culture assessment methods have been developed and applied to good effect. This paper explores some recent developments in the assessment of safety culture, drawing on UK defence and rail industry experience.

First it explores the case for defining the area of safety being assessed, such as occupational safety versus process safety. Experience shows that an organisation's safety attitudes and behaviours can vary from one aspect of safety to another. This leads to the question of whether safety culture assessments need to be specific to a defined area of safety or can assessments cover safety more generally. Lessons learnt from defence and rail indicate there may be circumstances where the area of safety needs to be defined and where assessment may need to be specific to an aspect of safety.

Secondly, it explores methods for assessing responses. Some methods present results, such as the average rating for a set of questions, to identify relative strengths and weaknesses. Other methods will also grade or categorise results using scales such as the Pathological to Generative scale. There are a number of competing categorical approaches which aim to enable results to be rated in absolute terms and provide a qualitative indication of safety culture maturity in an organisation. There are also examples of transforming questionnaires and other forms of assessment into categorical assessments of safety culture maturity.

Thirdly, whilst some methods focus on a single tool, such as employee questionnaires, others have moved towards a toolkit approach, such as using questionnaires and workshops. This reflects at least two considerations. First, doubts about the validity of assessments have raised the question of how best to boost the confidence that can be placed in results. This has led to the option of using a battery of methods and triangulating results, as a form of cross validation. Secondly, a battery of methods is sometimes advocated to fulfil the complementary goals of both measuring and understanding safety culture. A toolkit approach is being applied in the UK defence and rail sectors, offering lessons learnt concerning a range of safety culture assessment methods, their respective roles, advantages and disadvantages.

Keywords: Safety culture, assessment, attitudes and behaviours.

Defining the domain of safety

Divergent safety attitudes and behaviours

Experience shows that an organisation's safety attitudes and behaviours can vary from one aspect of safety to another. For example, in 2005, the Texas City oil refinery explosion killed 15 workers and injured more than 170 others, exemplifying how an organisation can excel in one aspect of safety (personal safety), yet fail so dramatically in another (process safety). Quotations from the Chemical Safety Board's (CSB) incident report (the 'Baker Report') included: "*BP has emphasized personal safety but not process safety,*" and "*...a very low personal injury rate at Texas City gave BP a misleading indicator of process safety performance*". BP was lulled into a false sense of security by becoming overly fixated with workplace safety performance.

Similarly, following passenger train derailments including Potters Bar (2003), Hatfield (2000) and Grayrigg (2009), the Office of Rail Regulation (ORR) commented on Network Rail's push for workplace safety, but acknowledged the comparatively poor focus on procedural compliance, timely completion and escalation of maintenance activity, and organisational understanding of their assets to determine the parameters of their safe operation.

These incidents indicate that attitudes towards different aspects of safety can vary within organisations, as illustrated in Figure 1. The figure suggests that the people may act differently in respect of, for example, occupational safety and process safety, in accordance with the attitudes prevailing in each safety domain. Therefore, a high rating for occupational safety attitudes and behaviours may or may not necessarily mean that the organisation will also have a high rating for process safety attitudes and behaviours, and vice versa.

Figure 1: Overlapping but distinct accident triangles
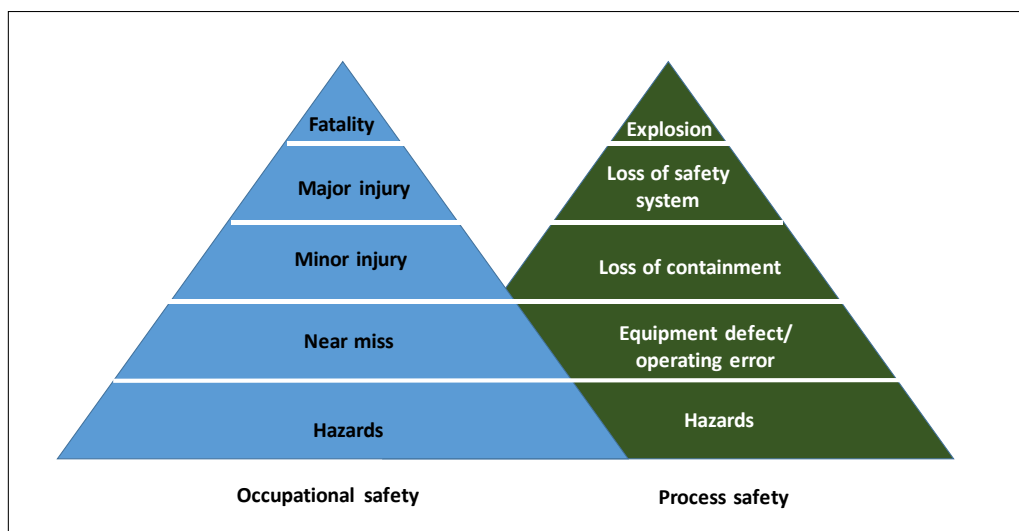
This highlights that attitudes and behaviours may be divergent across different areas of performance. In our own safety culture assessment experience, differences have also been found between occupational health and occupational safety, such as high levels of stress and fatigue being 'accepted' despite stringent occupational safety management. This is thought to be related to:

- Risk perceptions

  The extent to which an organisation recognises a hazard and perceives it to be a significant risk will influence the extent to which they manage it. Peoples' knowledge of hazards and risks, their experience of incidents, the level of regulation and stakeholder concerns interact to influence the perception of a hazard or safety domain.

  This can be compounded by competing business objectives differentially influencing areas of safety performance. For example, our assessment experience has found production pressures leading to maintenance 'workarounds' being commonly tolerated to meet operational schedules, whilst strictly enforcing personal safety of maintenance staff.

- Formalisation of policies for a safety domain

  Organisations may formalise policies and performance measures (such as Lost Time Injury rates) for a single area of safety, compounding skewed risk perceptions and subsequent behaviours.

- Compartmentalising safety domains

  Different aspects of safety are often managed by different departments (such as occupational safety departments versus engineering departments managing product safety), with different communities of safety professionals operating with different goals. This often arises from a need for domain specific expertise and domain specific safety procedures.

- Doman specific regulation

  There are many examples of domain specific regulation, such as for nuclear safety and major accident prevention, with specific requirements, such as for Major Accident Prevention Policies.

Implications for safety culture assessment

The latter observations do not necessarily mean that all organisations will have divergent or domain-specific safety attitudes and behaviours. An organisation may explicitly promote a holistic view of safety and clearly articulate that all aspects of safety should be equally managed. This may, and in our experience does, lead to people having a common view of all aspects of safety. In one example at a nuclear facility, we found that staff had a common view of radiological, nuclear and occupational safety. This could be related to the company promulgating a clear policy that all aspects of safety were of high importance and the adoption of a single all-embracing safety management system.

Accordingly, it needs to be clear which safety domains are being assessed in any assessment of safety culture. It should not be assumed that attitudes will be common across all areas of safety performance or that people will hold a common view of what "safety" comprises.

It should be noted that, where an organisation has domain specific safety arrangements, a similar level of maturity may be attained for each safety domain. It may also be that a higher standard is sought for one safety domain than another, in

accordance with the organisation's view of the relative risk associated with each safety domain. Therefore, having divergent or domain specific approaches to safety does not necessarily by itself impede safety performance. In our practical experience, safety culture maturity is influenced by the organisation's values, risk perceptions and safety goals, which might be common to all safety domains or domain specific.

The practical lessons for safety culture assessment are threefold. Firstly, preliminary work, such as a small-scale questionnaire pilot survey or a small number of workshops, can usefully explore how the organisation conceptualises safety. Is a singular concept espoused or are there multiple safety domains? If there is clear evidence that a singular concept is espoused, then a single measure of safety culture may be valid.

Secondly, preliminary work needs to explore what terms are used to refer to safety within the organisation. In our experience, organisations have used the following terms to refer to specific safety domains:

- Personal safety, workplace safety, radiological safety and occupational health and safety;
- Operational safety, fire safety, process safety, air safety, combat safety, marine safety, and nuclear safety;
- Product safety, design safety, food safety, and engineering safety;
- Patient safety and clinical risk.

Irrespective of whether there is a single or multiple safety domain, a clear definition of the area of safety being assessed should be stated, such as at the outset of any safety questionnaire or other assessment tool, ideally with examples of typical safety policy and practices. This helps ensure that respondents are thinking of the same safety area(s) of safety performance when answering questions. Failure to define the area of safety being assessed creates a risk of non-comparable responses from across an organisation and uncertainty over what aspect(s) of safety performance is being measured.

Thirdly, where there is evidence of multiple safety domains, this may necessitate completing independent if parallel assessments. In our work with BAE Systems (Lockwood et al., 2015), a comparable assessment of Product Safety and of Workplace safety is completed in parallel, using the same tools and matched question sets. The terms applied in each tool are specific to the safety domain (e.g. a Workplace Safety questionnaire and a Product Safety Questionnaire) and supported by definitions and exemplification of the safety domains. An alternative is to limit an assessment to a single safety domain, such as process safety. This is similar to common practice in some other sectors, where assessment may be specific to, for example, nuclear safety (Morrow, 2012).

Results from safety culture assessments

Do we find divergent responses? In practice, our assessments have found that attitudes can be markedly different, with, on occasion, one safety domain being managed to a higher standard than another. This can relate to the business imperatives and business performance measures in effect, leading to behaviours diverging.

In some cases, the measured level of safety culture maturity may be similar across safety domains, but it can be found (through for example workshops) that the behavioural drivers differ and each safety domain is perceived discretely. The common level of maturity can hide a divergent approach to managing each safety domain. This may lead to a need to adopt different approaches to improving each area of safety, due to people perceiving different drivers and objectives. For example, in one case staff indicated that their ability to engage in improving the safety of a manufactured product was limited by their product knowledge, something which did not limit their engagement in occupational safety.

Assessing and interpreting results

Reviewing options

A key issue concerns the interpretation of safety culture assessment results. Some staff questionnaires provide results in terms of the proportion of staff who agree or strongly agree that there is, for example, effective safety leadership. This is useful with respect to identifying relative strengths and weaknesses. However, there is also a need to assess the level of maturity in absolute terms, and to be able to compare results to other similar organisations. For example, if 70% of staff agree there is effective safety leadership, is this 'good enough'? Does it indicate that it is *essential* to improve leadership and invest significant time and resource or does a result of 70% indicate that leadership is 'adequate' and attention should focus on other aspects of culture? Acquiring safety culture results for other organisations is, generally, impractical. In addition, a wide range of safety culture methods are used, with different measurement methods and question sets that cannot be directly compared.

Categorical assessment

Some approaches involve categorising safety culture maturity. For example, the Keil Centre Safety Culture Maturity Model, EI Hearts and Minds Tools, Manchester Patient Safety Framework & the MAA Safety Culture Evaluation Tool, categorise safety culture scores by maturity. The exact approach differs between these methods, with some including measurement (categorised) and others relying on qualitative categorical assessment. However, in each case an organisation is 'benchmarked' against a safety culture maturity scale to help them understand how 'good' or 'poor' the result is. This highlights the option of using a categorical (ordinal) assessment scale for grading safety culture maturity.

Our option is to use the Pathological to Generative scale adapted by Hudson (1999), which is used within the Hearts and Mind toolkit (and some other methods). This scale is advocated due to it a) being an open source, b) a progressive model, c) widely applied and d) intuitive, with the scale points supported by safety culture research.

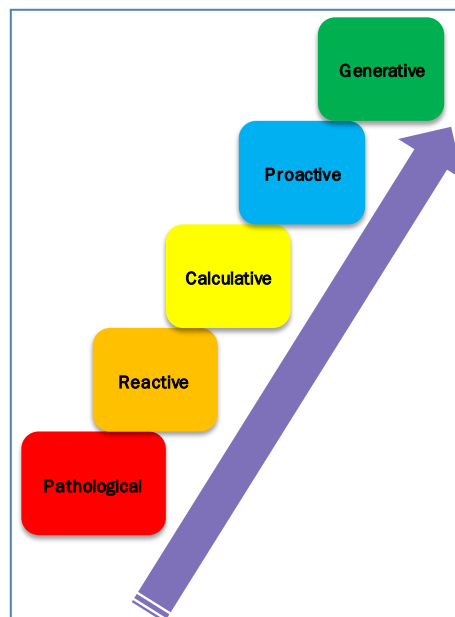There are examples of questionnaire survey scores being assessed using categorical scales, as per Figure 3

Figure 3. A key rationale underpinning these scales is that safety critical and major hazard operators require a very high level of maturity for their culture to be considered 'suitable and sufficient'. A rhetorical question is whether it would be considered 'good' if (for example) 30% of staff reported that operating procedures were impractical? The example scales in Figure 3 assume that results such as 50% agreeing that leadership is effective, is poor', rather than 'average', whilst requiring a very high score to be categorised as 'good'.

Figure 2: Generative to pathological culture maturity scale

Figure 3: Examples of categorical assessment of safety culture results.

NOSACQ-50		Caterpillar Safety Perception Survey		CAA ShoMe Tool	
Typology	% Score of maximum	Rating	% Score of maximum	Rating	% Score of maximum
Generative	≥82%	Good	≥90%	Good	≥85%
Proactive	75-81%				
Calculative	68% to 74%	Average	75% to 89%	Average	70% to 84%
Reactive	59-67%			Poor	50 -69%
Pathological	<59%	Poor	<75%	V Poor	<50%

A preferred categorical scale

These considerations were combined in our work with BAE Systems to develop a scale, based on the Pathological to Generative scale, as per Figure 4. The scale was thought to enable an absolute assessment to be made of the level of safety culture maturity and comparison with the standards expected of comparable safety critical organisations. Whilst there is no definitive guidance on what standard of culture a safety critical organisation should aim for, experience suggests that a 'proactive' culture that aspires towards continuous improvement and prevention of risks (rather than only seeking to fulfil specific safety management duties) is consistent with common expectations for safety critical / high hazard organisations.

The scale is applicable to quantitative questionnaire and qualitative forms of assessment, such as workshops. Each point on the scale was defined in qualitative terms – elaborating the types of attitudes and behaviours per point on the scale. Questionnaire results from 5 point Likert scales can be translated into an average out of 1 to 5. This allows qualitative and quantitative forms of measurement to be assessed using the same scale.

Figure 4: Categorical assessment scale

Scale	Score (1 to 5)
Generative	**≥4.5**
Proactive	**4 to 4.49**
Managing	3.5 to 3.99
Reactive	3 to 3.49
Dysfunctional (pathological)	**<3**

Common elements of safety culture

Another key question, is what elements of safety culture should be assessed? Therefore, safety culture assessment tools, guides and frameworks were reviewed to identify a common set of safety culture elements measured and/or described by these instruments (Lockwood et al., 2015). A qualitative approach was taken to reviewing the instruments. This involved:

- Detailing the safety culture elements measured and/or described by each tool, guide or framework;

- Reviewing element structures;

- Labelling the common elements based on the underpinning topic of each element and;

- Finalising an identified set of common elements of safety culture.

The resulting elements are:

- Leadership and commitment;

- Communications;

- Ownership and accountability;

- Balancing safety and other responsibilities;

- Perceptions of risks and risk controls

- Commitment to training and competence;

- Teamwork and relationships;

- Combating normalisation of deviance;

- Open challenging culture;

- Learning and sharing culture.

These elements were tested and consulted on, including a factor analysis and a reliability analysis (Cronbach alpha) of responses to a safety culture questionnaire survey. The results were highly supportive of one overarching attitude towards safety among the respondents that informs their views on product and workplace safety respectively, with most questions loading within a single factor. This led to a condensed set of safety culture questions.

It was found that the factor structure is 'unstable', changing somewhat from one assessment to another. Thus, whilst statistical analysis may point to a similar set of questions and elements, experience indicates that there is no single 'definitive' set of safety culture elements. The lesson from this being that the element structure of safety culture assessment tools cannot be determined solely from statistical methods but needs to also be informed by qualitative research and models of safety culture.

Multi method assessment

Thirdly, whilst some assessment methods focus on a single tool, such as employee questionnaires, others (such as the RSSB toolkit) have moved towards a toolkit approach, such as using questionnaires and workshops. This reflects at least two considerations. First, doubts about the validity of assessments have raised the question of how best to boost the confidence that can be placed in results. This has led to the option of using a battery of methods and triangulating results (see Figure 5), as a form of cross validation. Secondly, a battery of methods is sometimes advocated to fulfil the complementary goals of both measuring and understanding safety culture. The aforementioned work by BAE Systems (Lockwood et al., 2015) involves the use of:

Figure 5: Triangulating assessment results

- Safety culture staff questionnaires;

- Perceptions workshops, exploring staff perceptions of attitudes and behaviours, the drivers behind behaviours and examples of behaviours;

- Observation of behaviours and communications, such as during business meetings, toolbox talks and whilst people are at work;

- Review of safety documentation, such as policy and procedures;

- Review of leading and lagging safety culture indicators, such as rates of procedural non-compliance and proportions of incidents involving unsafe behaviour;

- Critical incident analysis using root cause analysis to explore actual attitudes and behaviours that contributed to incidents and the response to these incidents;

- Focus groups - to help understand the reasons for observed and perceived attitudes and behaviours.
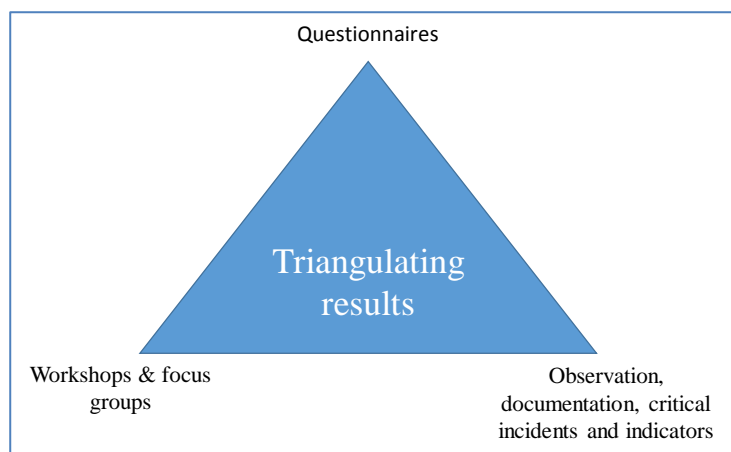
All forms of assessment can offer value to an assessment. For example, questionnaires are a cost effective method for measuring culture within and across departments of large organisations, and enable comparison of results from one year to another. However, questionnaires tend not to provide in depth qualitative understanding of attitudes and behaviours for which perceptions workshops are better suited. Workshops and focus groups enable in depth exploration of perceptions, drivers and behaviours, but due to practical constraints tend to involve a relatively small number of persons. Whilst observations of behaviour can be powerful, practical constraints can limit the extent of observation and observed behaviours may not reveal underpinning attitudes. Critical incidents can also be powerful examples of actual behaviour but in practice significant incidents can be infrequent or not recent enough to the representative of current behaviour. Leading and lagging safety culture indicators can provide powerful measures of behaviour but are often unavailable or do not cover the full range of attitudes and behaviours. Safety documentation can provide an indication of management intent but requires verification of implementation. In practice, a business may choose to apply a selection of methods, such as questionnaires, workshops and focus groups, depending on business specific requirements and the availability of, for example, incidents to review. However, good practice and practical experience recommend the use of a combination of questionnaires, workshops and focus groups as a minimum, in order to measure and understand the safety culture of an organisation.

In each case, a set of safety culture questions have been developed that align to the Pathological (re-termed Dysfunctional for BAE Systems) to Generative scale and a common set of safety culture elements. This allows results from each method to be directly compared and assessed. The comparison of results also helps provide a greater weight of 'evidence'. If the results from each form of assessment are consistent, this gives greater confidence in the assessment results. In practice we have found, on occasion, that results may differ between forms of assessment. This has been found to be due to:

- Observations of behaviour potentially being skewed to observable words and actions, which may reflect formal safety management processes. This tends to lead to a rating of 'managing' culture;

- Critical incident analysis potentially being skewed towards examples of 'poor' behaviour leading to lower ratings of culture;

- People offering higher ratings within workshops than justified by the examples of behaviour cited in their support, possibly due to a tendency towards 'socially desirable responses';

- Questionnaire responses varying across departments and roles.

This leads to a need to verify the representativeness of responses from each form of assessment. If an assessment is initially skewed, a further round of assessments may be needed in order to achieve a more representative set of results. For example, examples of positive 'incidents' may be sought out to balance examples of adverse incidents, or a wider range of activities may be observed if initial observations are overly limited in scope.

On occasion, the results from workshops may need to be re-rated or weighted to match the examples of behaviour cited by delegates. For examples, workshop delegates can rate the organisation as 'Proactive' whilst citing examples of 'Reactive' management behaviour. Similarly, if behavioural observation or critical incident analysis is found to be skewed or unrepresentative, the results from these forms of assessment may need to be awarded less weight. Where results differ by department or role, then a single overall result for a business may be misleading, with results instead being only presented by department and role. In all these cases, in practice, the assessors need to understand what has been assessed, how representative the results are and the reasons for any skews or bias in results, with the weight awarded within conclusions to each form of assessment adjusted by the confidence placed in each assessment method.

A common issue across all forms of assessment is the extent to which they provide a representative and valid measure. It is equally important to have a sufficient response rate to a questionnaire as it is to have sufficient attendance at workshops. In our experience, this relies on:

- The extent to which the senior management of a business expresses a wish for management and staff to commit time to the assessment;

- Senior management commitment being cascaded to managers, supervisors and staff through face to face communication, engaged with staff representatives, as well as media such as email and newsletters;

- The extent to which the assessment process, timetable and requirements are communicated to managers and staff;

- Matching the specific process of assessment to the needs of staff, such as having paper questionnaires for operators versus online questionnaires for office based staff;

- The extent to which participation in an assessment and its representativeness is tracked and additional action taken to boost low response rates and secure more representative assessments.

The adoption of an effective communication and implementation process can achieve high response rates in the region of 70% or more, and high participation in workshops, for example. Where reliance is placed on passive forms of communication, such as email, this can be associated with low response rates in the region of 15% and minimal workshop participation which can undermine confidence in the results.

Lessons learnt

Safety culture assessment is commonly completed by safety critical and high hazard organisations, with results used to inform safety improvement programmes. The research and practical experience summarised here offers a number of key lessons, including:

- Verifying whether a single measure of safety should be pursued or whether one or more domain specific assessments of safety culture are required;

- Using a categorical form of assessment across all forms of assessment to allow results to be assessed in absolute terms, i.e. what is the level of safety culture maturity, and to allow results to be compared across businesses, departments and roles;

- Using a multi method assessment strategy to enable cross validation of results and consideration of the level of confidence in results;

- Using qualitative and quantitative forms of assessment to provide an in depth understanding of the drivers for culture as well as a measure of maturity;

- Developing a qualitative understanding of the representativeness and validity of results from each form of assessment, to allow differences between methods in results to be understood and to allow results from each form of assessment to be weighted;

- To implement an effective communication, tracking and follow up process to achieve a high and representative level of participation in the assessment.

A well communicated and supported assessment can by itself support safety performance through the demonstration of organisational commitment and engagement with staff, as well as achieving a powerful set of results. Conversely, a poorly communicated and under supported assessment can cause staff cynicism and adverse perceptions of organisational safety commitment. Where the results of an assessment are shared with staff along with a clearly communicated improvement plan, this can by itself improve staff perceptions and confidence in organisational commitment to safety. Where results are not shared and little action arises from an assessment, this too can cause adverse staff perceptions. Accordingly, an assessment

of safety culture should ideally only be embarked upon when it is clear that senior management will effectively support it and follow up on the results. The assessment of safety culture is not an abstract measurement activity. The act of carrying out an assessment can affect staff perceptions for better or worse, whilst also providing a rich body of information on which to devise further improvements to safety.

References

Baker, J., Bowman, F., Erwin, G., Gorton, S., Hendershot, D., Leveson, N., Priest, S., Rosanthal, I., Tebo, P., Wiegmann, D., and Wilson, L. The BP U.S. Refineries independent safety review panel. January 2007. http://www.csb.gov/assets/1/19/Baker_panel_report1.pdf . Accessed December 2016.

Caterpillar (2011) The Safety Perception Survey. http://www.cat.com/en_US/support/safety/safety-culture-consulting/safety-perception-survey.html Accessed December 2016.

Civil Aviation Authority (CAA) (2003). Introduction to the Safety Health of Maintenance Engineering (SHoMe) Tool. CAA paper 2003/12.

Energy Institute (2011). Process Safety Survey. Retrieved from: https://www.energyinst.org/technical/PSM/process-safety-survey. Accessed December 2016.

Eurocontrol (2009). ATM Safety Framework Maturity Survey Methodology for ANSPs. Document reference: ESP-2009-78.

Hudson, P. Paper presented at the RTO HEM Workshop on "The Human Factor in System Reliability - Is Human Performance Predictable? " Held in Siena, Italy, 1-2 December 1999, and published in RIO MP-032.

Kines, P., Lappalainen, J., Mikkelsen, K. L., Olsen, E., Pousette, A., Tharaldsen, J., Tómasson, K. & Törner, M. (2011). Nordic Safety Climate Questionnaire (NOSACQ-50): a new tool for measuring occupational safety climate. *International Journal of Industrial Ergonomics, 41*, p. 634-646.

Lockwood, F., Canham, R. & Wright, M. (2015) Developing a framework for assessing Workplace and Product Safety culture in BAE Systems. In Proceedings volume of the 2015 Safety-critical System Symposium (SSS'15). http://scsc.org.uk/paper_129/Lockwood - Developing a Framework for Assessing Workplace and Product Safety Culture in BAE Systems.pdf?pap=980 Accessed December 2016.

Morrow, S. (2012). *Independent Evaluation of INPO's Nuclear Safety Culture Survey and Construct Validation Study*. Retrieved from http://pbadupws.nrc.gov/docs/ML1217/ML12172A093.pdf Accessed 1 Apr. 2016.

Office of Rail Regulation. Train Derailment at Hatfield: A Final Report by the Independent Investigation Board, July 2006. http://webarchive.nationalarchives.gov.uk/20131001175041/http://www.rail-reg.gov.uk/upload/pdf/297.pdf Accessed December 2016.

Rail Accident Investigation Branch, Department for Transport. Rail Accident Report. Derailment at Grayrigg 23 February 2007. Report 20/2008 v5 July 2011 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/411798/081023_R202008_Grayrigg_v5.pdf Accessed December 2016.

Rail Accident Investigation Branch, Department for Transport. Rail Accident Report. Derailment at Grayrigg, Cumbria 23 February 2007. Report IR2/2007. October 2007. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/528168/IR22007_071003_Grayrigg.pdf Accessed December 2016.

Rail Safety and Standards Board. Safety Culture Toolkit. Retrieved from: http://safetyculturetoolkit.rssb.co.uk/home.aspx. Accessed December 2016.