
Critical review of safety performance metrics

Nektarios Karanikas

Faculty of Technology/Aviation Academy,
Amsterdam University of Applied Sciences,
Weesperzijde 190, 1097 DZ Amsterdam, Netherlands
Email: nektkar@gmail.com
Email: n.karanikas@hva.nl

Abstract: Various tools for safety performance measurement have been introduced in order to fulfil the need for safety monitoring in organisations, which is tightly related to their overall performance and achievement of their business goals. Such tools include accident rates, benchmarking, safety culture and climate assessments, cost-effectiveness studies, etc. The current work reviews the most representative methods for safety performance evaluation that have been suggested and applied by a variety of organisations, safety authorities and agencies. This paper discusses several viewpoints of the applicability, feasibility and appropriateness of such tools, based on the viewpoints of managers and safety experts involved in a relevant research that was conducted in a large aviation organisation. The extensive literature cited, the discussion topics, along with the conclusions and recommendations derived, might be considered by any organisation that seeks a realistic safety performance assessment and establishment of effective measurement tools.

Keywords: safety performance; safety metrics; safety assessment; safety audits; performance indicators; safety culture; safety climate.

Reference to this paper should be made as follows: Karanikas, N. (xxxx) 'Critical review of safety performance metrics', *Int. J. Business Performance Management*, Vol. X, No. Y, pp.xxx-xxx.

Biographical notes: Nektarios Karanikas is an Associate Professor (Hoofddocent) Safety and Human Factors in the Aviation Academy of the Amsterdam University of Applied Sciences. He studied aircraft engineering in the Hellenic Air Force Academy. He received his MSc Human Factors and Safety Assessment in Aeronautics at Cranfield University, and he was awarded a Doctorate in Safety and Quality Management from Middlesex University. He has completed a variety of professional courses (e.g., safety officers, aviation safety management, operations management, mishap investigation and failure analysis courses), he holds the CEng, GradIOSH and PMP professional certifications and he is member of various European and international bodies (HFES European Charter, FSF, EAAP, ISASI, IET and PMI).

1 Introduction

The assessment of safety performance comprises valuable means for managers towards implementation of best practices, thus leading to the highest probable safety outcome. Safety performance metrics are essential for directors, senior and line managers,

supervisors and safety professionals, who need a tool for monitoring goal achievements, progress, and expected results. As Arezes and Miguel (2003), Channing and Ridley (2008), Easter et al. (2004) and Stranks (2006, 2008) stated, such metrics should be consistent with the overall organisational performance. International Civil Aviation Organization (ICAO, 2013) linked safety performance with organisational safety health, the latter including the exploration of poor safety symptoms, adoption of acceptable risk levels, conduction of inspections, surveys and audits, and requirement for a quality assurance system regarding safety performance assessment.

However, the establishment of safety performance metrics does not always guarantee a timely response to critical safety issues. For example, in their discussion about their experience regarding the Texas refinery accident in 2005, Boyle et al. (2010) stated that although British Petroleum Company had established numerous safety indicators, these did not include leading ones that could drive actions prior to the accident. The authors suggested that the organisations establish safety metrics, which will provide information on the risk mitigation effectiveness. These measures might be proactive (e.g., monitoring of hazardous cases before causing an accident or incident), reactive (e.g., measurement of factors that have already contributed to safety events), and driven by external factors (e.g., knowledge and expertise from other industries). Similarly, the Transport Canada (TC, 2002, 2004) proposed a reactive and proactive monitoring of safety performance that will allow achievement of safety goals; the monitoring results are highly valued in improving safety management systems (SMS).

Taking into consideration the aforementioned concepts, the current study firstly reviewed the various performance measurement tools as safety authorities and agencies, experts and practitioners have suggested, thus presenting extended information regarding safety assessment methods. This review is followed by a discussion based on a search conducted in regards to safety performance assessment in a large aviation organisation, which has been under transition from a safety program to an SMS approach. The specific research employed various professional viewpoints from managers with rich operational experience and/or expertise in safety programs development and management. The conclusions and recommendations are deemed useful for organisations, which emphasise in their effective safety performance evaluation, or safety authorities and agencies that seek to establish realistic and indicative safety performance metrics.

2 Literature review

2.1 Basics of performance measurement

Kaufman (2011) presented two major models of performance improvement: The reactive one, based on engineering design and comprising a reaction to the deficiencies indicated during operations, and the proactive model based on creative thinking and planning prior of any unwanted event. The latter approach was more favourable to the author, who advised performance professionals to adopt a more holistic performance assessment approach and challenge any organisational visions and missions that have been developed based on reactivity. In his review on safety performance literature and practice, Parker (2000) acknowledged that almost every organisation measures performance in order to identify success, explore if it meets customer requirements, reveal potential weak internal

processes, support its decisions with facts, and monitor the implementation of improvements. The author discussed various measurement methods:

- financial measurements (e.g., rate of return)
- benchmarking, as a tool for comparisons among similar units
- the balance scorecard, which is used to quantify the business goals and strategy by introducing measures around several perspectives (financial, customer-oriented, internal processes, innovation, and improvement)
- the activity-based costing, that attributes the cost to activities and products more accurately than traditional accounting methods.

Moreover, Parker (2000) distinguished the metrics in lagging and leading, and in outcome, action, input and diagnostic ones. The author suggested the following fundamentals of performance measurement:

- alignment of performance metrics with organisations' strategy
- inclusion of every sub-unit in the organisation-wide metrics
- management commitment to the measurement regime
- performance improvement through measurement
- reliability of performance assessment methods.

In Sagar et al. (2013), following their literature review regarding the development of performance measurement methods from 1991 to 2011, concluded that although plenty of performance measurement approaches have been suggested in the literature, a large number of them have not been practically tested. The authors also proposed that organisations moved beyond the use of conventional score cards, to measures that are holistic, integrated, dynamic and effective. As a result of another relevant extensive literature review, Bititci et al. (2012) noticed that performance measurement has been developed in alignment with the progress indicated in the business trends, moving from mere productivity control to budgetary control and, finally, to integrated performance management. As the authors claimed, the future challenge is to understand performance measurement as a social and learning system.

2.2 Safety performance measurement approaches

In their discussion about safety performance evaluation tools, Adebisi et al. (2007) proposed the following methods:

- a statistical approach, which employs correlations, regressions, non-parametric statistics and variance analysis and results to figures such as accident frequency and severity, time delays, etc.
- probability studies of expected accidents based on actual accident numbers
- risk assessment approaches

- control charts that are based on random sampling, depict the distance of activities from safety standards, and lead to measures' implementation whenever the thresholds are exceeded
- comparisons amongst annual accident costs and accident prevention capital invested
- questionnaire surveys about safety initiatives effects on accident prevention in the workplace.

In addition, Adebisi et al. (2007) discussed various methods for quantifying safety performance:

- accident rates, including frequency of injuries
- efficiency of the safety program with estimations that include accident costs and resource investment on safety for specific periods
- expected number of safe work activities based on accident occurrence
- productivity of a safety program by employing quantitative and qualitative measurements of resources and accidents.

However, as the aforementioned authors discussed, although such models and methods may be useful for getting a superficial and initial picture of safety levels, these:

- Do not address management problems.
- Focus on failures rather than causes.
- Occasionally lack validity because accident implications do not regard only money costs. Accidents inevitably cause adverse psychological, social, professional and organisational effects that cannot be quantified.

Stranks (2006, 2008) in his examples of safety monitoring data included:

- the extent to which objectives and targets have been set and met
- employees' perception of management commitment to safety
- adequacy of communication concerning safety policy and documentation
- the extent of compliance to legal and international standards
- percentage of risk assessments carried against the ones planned
- time required to implement safety-related remedial actions
- frequency of monitoring activities (audits, surveys, inspections, etc.).

ICAO (2013) and Goglia et al. (2008) suggested gap analysis as a requirement to compare the safety arrangements of an organisation with the ones that are necessary for the functionality of a complete SMS. Moving beyond compliance, ICAO (2013) stated the need to establish performance-based safety assessment methods, as means to evaluate the effectiveness of SMS. Under this concept, the European Aviation Safety Agency (EASA, 2014) concluded the need to complement the SMS prescriptive regulatory framework with a performance-based approach. Some initiatives towards assessment of safety performance in the aviation context include:

- the SMS evaluation tool developed by the Safety Management International Collaboration Group (SMICG, 2012), which addresses characteristics of presence, suitability, operability and effectiveness of a broad spectrum of SMS activities.
- the effectiveness of safety management (EoSM) instrument, which was devised by the Eurocontrol (2012) and assesses the maturity of SMS in a five-stage classification: initiating, planning, implementing, managing and measuring, and continuous improvement.

Barrie (1990) viewed measurement and assessment as vital for any organisation that attempts to quantify and monitor management goals; the author proposed the following fundamental characteristics of an acceptable and objective measurement regime:

- Management processes determine the form and nature of the measurements.
- The numerators and denominators of any ratio shall refer exactly to the same conditions and factors.
- The indicators referring to a specific activity must be under control by the employees serving this activity.
- Since the decisions often need considerable time to cause changes and results, the reference period of any indicator adapted must be considered.
- High-level measurements must be based on lower level measures, the same way that organisational goals are the outcome of several sub-goals.
- Long-term indicators must be favoured against short-term ones, due to daily and seasonal variability of conditions and individuals.
- The measures must be consistent and demonstrate validity and reliability in order to be applied over time, regardless of the person(s) calculating the indicators.

A theoretical model derived from 200 articles on job performance, safety performance, safety climate and safety training, was introduced and tested by Burke et al. (2002), who proposed a four-factor safety measurement evaluation:

- usage of personal protective equipment
- engagement of employees in risk reduction
- communication of health and safety information
- exercise of employees' rights and responsibilities.

The normalisation of safety related data might be achieved by the usage of rates or frequencies, categorisation in groups, selection of specific projects for comparison, establishment of relative weighting factors, and comparison of results with a mathematical model derived from data (Stapenhurst, 2009).

2.3 Safety performance indicators

ICAO (2013) stated that in order to assess safety performance, the organisation must decide its safety indicators and targets along with their quantification models. The selection of indicators depends on the level of safety to be represented, either generic or

specific, and must be driven by their representativeness of the outcomes, processes and functions of the level under consideration. ICAO (2013) also distinguished between safety measurement and safety performance: The former referring to high-level and high-consequences safety events, and the latter including low-level operational processes.

Moreover, the TC (2002, 2004) acknowledged that accident rates might be effective as a reactive performance indicator only when they are constantly high; in any other cases, they might lead to a false impression that nil accidents associate with high safety level, driving organisations to fail in recognising latent conditions. Stranks (1994) criticised the use of accident data in safety performance measurement and argued that any management that relies on safety audits to ensure compliance to legislation, and monitors safety performance only through accident rates, could not claim competence and commitment to safety. According to the author:

- accident rates measure failure, not success, and they present random fluctuations
- there are time delays between safety measures' introduction and their actual implementation
- occupational diseases are not adequately measured
- there is an emphasis on the actual accident severity, and not on their potential for re-occurrence
- safety events are subject to under-reporting
- low accident rates do not contribute to predictions of future accidents.

Table 1 Key performance indicators

<i>Category</i>	<i>KPI</i>
Regulatory responsibility	Understanding of regulatory responsibilities
Risk control	Identification of hazards
	Safety procedures
	Safeguards
	Assessment of training needs
	Health control
Enabling activities	Willingness to use external H&S information and support
	Workforce involvement/participation
	Communication of safety information to the workforce
	Incident/accident investigation

Source: Quoted from Amey Vectra (2000)

The research of Amey Vectra (2000) resulted in the health and safety key performance indicators (KPI) shown in Table 1, which were scored according to the scale shown in Table 2. The assessment approach is similar to the auditing tools discussed below, accompanied with questions to managers under the scope of revealing safety management policies and their implementation. Al-Homoud and Khan (2004) followed a similar methodology for the assessment of safety measures in residential buildings in Saudi Arabia.

Table 2 Scoring system for KPI

<i>Score</i>	<i>Meaning</i>
5	Good system and used
4	Reasonable system and used
3	Partial system and used
2	System not effectively used
1	Poor system

Source: Quoted from Amey Vectra (2000)

Finally, an exploration of safety professionals' viewpoints about modelling SMS with software unveiled that most of the safety performance indicators employed include (ASMSG, 2012):

- average duration needed to close out issues derived from accident-incident reports, safety audits, safety meetings, hazard reports, etc.
- number of safety related reports submitted through safety reporting systems
- number of safety meetings and attendants
- costs of safety events, accidents and incidents
- weighing of performance indicators in terms of their contribution to safety.

2.4 Safety audits and benchmarking

The TC (2005) suggested audits in the form of site visits and observations as basic tools for safety management assessment; the specific authority introduced a scoring system mainly based on the conformity of the audited SMS to the described components and elements (Table 3). Similar models for safety performance evaluation are the International Safety Rating System, the European Foundation for Quality Management, the American Volunteer Protection Program, and the Canadian Partners for Injury Reduction Health and Safety Audit, cited in Gholami (2011).

Cameron and Duff (2007) in their research on construction context evaluated positively the use of an audit instrument that measures safety management behaviour. Based on interviews held with employees, they developed a tool that uses a three-scale measurement level to assess the areas of induction training, toolbox training, safety committees, subcontractors' safety, safety records, safety manager actions, and safety considerations. Downs (2003) proposed a method named Safety and Environmental Management System Assessment (SEMSA), which is based on a formal process alternative to auditing and emphasises on five management categories: management's expectations and communication, risk assessment and action plans, implementation of processes, checking and corrective actions, and review-renewal.

Table 3 Transport Canada SMS assessment components and elements

<i>Table A – SMS assessment protocol framework</i>			
<i>Component</i>		<i>Element</i>	
0	Safety management system		
1	Safety management plan	1.1	Safety policy
		1.2	Non-punitive safety reporting policy
		1.3	Roles, responsibilities and employee involvement
		1.4	Communication
		1.5	Safety planning, objectives and goals
		1.6	Performance measurement
		1.7	Management review
2	Documentation	2.1	Identification and maintenance of application regulations
		2.2	SMS documentation
		2.3	Records management
3	Safety oversight	3.1	Reactive process
		3.2	Proactive process
		3.3	Investigation and analysis
		3.4	Risk management
4	Training	4.1	Training, awareness and competence
5	Quality assessment	5.1	Operational quality assurance
6	Emergency preparedness	6.1	Emergency preparedness and response

Source: Quoted from TC, 2005)

Safety performance benchmarking specifically through audits, as presented by Fuller (1999), revealed that the assessment of basic health and safety elements (i.e., policy, organisation, planning, measurement, audit and review), may simultaneously comprise a basis for establishing performance measurements with the proper use of rates and scales. Also, Stranks (1994, 2006, 2008), Ferrett and Hughes (2007), and Boyle (2008) presented benchmarking as a quality management tool towards improvement when accompanied by actions plans that are specific, measurable, agreed, realistic, traceable and time bound (SMARTT). Regardless of the method of measurement, several criteria must apply:

- the measurement system must demonstrate practicality
- the measures must be sensitive enough to capture changes
- the technique must demonstrate reliability, stability, validity, objectivity and accuracy
- the measurement must be efficient and understandable.

Similarly to benchmarking, Fuller (1997) introduced a safety performance assessment method based on intra-company and inter-company audits; safety management areas were divided in several subsections that were assessed either during local audits regarding the intra-company part, or by each safety manager regarding the inter-company benchmarking. The main measurement areas were: Policy, organisation, planning, risk

management, performance and auditing. The research outcomes identified strong correlations of topics – areas measured with the overall safety performance and suggested the use of the particular method as a tool for monitoring the effects of each individual topic to the overall safety level.

2.5 Safety cost-effectiveness

Marlow et al. (2004) discussed thoroughly safety performance and its economic facets and introduced a software model named productivity assessment tool. The authors presented in detail the following fundamental approaches on economic analysis models:

- return on investment (ROI) analysis, leading to the estimation of costs pay-off for a predetermined period
- cost-effectiveness analysis for estimating costs for each unit ‘saved’ that cannot be priced (e.g., human life)
- cost-benefit analysis that, in addition to direct costs, attempts to quantify all social and cultural aspects, which are often important but ignored.

Gitelman et al. (2008) assessed road safety measures based on the cost effectiveness of safety measures-based on estimation of accidents prevented, and a benefit-cost ratio of such measures according to the investment required. The study used mostly ‘hard’ accident costs (e.g., medical costs, lost productive capacity, administrative costs, property damage costs, loss of welfare) and focused solely on the local accident implications; the side effects of professional, social and organisational effects were not included. Under the same concept, Moses and Savage (1997) presented a cost-benefit analysis of the USA motor carrier safety programs, which showed that:

- safety audits and roadside inspections alone brought correspondingly 7% annual and 3% reduction in accident rates in a three months period
- audits presented a cost-benefit ratio of 1:4
- inspections’ cost-benefit ratio was about 1:1.5.

2.6 Safety culture and climate assessment

Beyond reactive performance indicators (e.g., accident rates), Arezes and Miguel (2003) focused on safety culture assessment as an important method, which in combination with the traditional rate measurements may reveal the ‘soft’ safety issues (i.e., psychological, social, professional and individual factors). According to the aforementioned authors, such a tool must refer to training and competence, job security and satisfaction, job pressure, communication, management commitment to safety, accident investigations, involvement in safety initiatives, errors and violations, and the subjective perception of safety culture inside the organisation under research.

In the same spirit, Mitchell et al. (2002) and Ferrett and Hughes (2007, 2011) discussed the usefulness of safety culture assessment instruments, both qualitative and quantitative ones, based on the fact that such measures may reveal positive or negative aspects of the upper-management commitment to safety, management involvement, employee empowerment, rewarding systems and reporting systems. On the opposite side,

Davies et al. (2003) claimed that any attempt to measure safety culture through questionnaires along with the definition of safety culture as the response to these questionnaires, leads to a circular system with no external referent; hence, 'safety culture' cannot be seen as an entity outside the questionnaire itself.

Kines et al. (2011) developed and tested the Nordic Climate Questionnaire as a tool for safety climate assessment. The questionnaire measures various components of safety management: priority, commitment, competence, empowerment, justice, risk control, communication, learning and trust. Simon (2005) introduced a culture assessment tool, named Simon open system culture change model, which is based on the perceptions, interviews and observations regarding organisational and cultural influences, as depicted in Table 4.

Table 4 Itemisation and definition of organisational and culture influences

<i>Organisational influences</i>	<i>Definition</i>
Technology	How the work is done.
Program Structure	Training, policy, procedure, etc.
Rewards	Promotions, compensation, awards.
Measurements	Leading as well as lagging indicators of safety performance
Social Processes	Trust, communication, caring, relationships
Environment	External business pressures to improve safety performance such as government regulations, customers, stockholders, workers' compensation costs, and the market place.
<i>Cultural influences</i>	
Leadership	Establishes vision and sets example for the new safety culture in a way that leads the organisation towards zero injuries.
Symbols	Physical or visual reminders of important safety values.
Values	Spoken principles such as 'people are more important than numbers' that guide the decisions of workers and managers.
Heroes	Organisational members that role model the values.
Rituals	Regular celebrations, ceremonies or activities that reinforce the importance of safety.
Norms and assumptions	Norms are the group's expectations for safety behaviour. Assumptions are the beliefs about what is safe or unsafe and why it is commonly accepted to perform a job in a safe or unsafe manner.

Source: Adapted from Simon (2005)

Gibbons and Thaden (2008) suggested the safety culture indicator scale measurement system (SCISMS); the authors invited respondents to use a Likert-scale for rating dimensions of safety culture on a variety of scales. These scales were derived from the literature referring to the main aspects of safety management and were adapted appropriately for pilots and aviation maintenance personnel (Tables 5 and 6).

Table 5 Scale inventory for flight operations version of the SCISMS

<i>SCISMS major factor scale</i>	<i>Sub factor scales</i>
Organisation commitment	Safety values
	Safety fundamentals
	Going beyond compliance
Operations interactions	Chief/fleet pilots
	Instructors/training
	Dispatch
	Operations control
	Ground handling/ramp operations
	Maintenance/engineering
	Cabin crew
Formal safety indicators	Reporting system
	Response and feedback
	Safety personnel
Informal safety indicators	Accountability
	Pilot's authority
	Professionalism

Source: Quoted from Gibbons and Thaden (2008)

Table 6 Scale inventory for maintenance operations version of the SCISMS

<i>SCISMS major factor scale</i>	<i>Sub factor scales</i>
Organisation commitment	Safety values
	Safety fundamentals
	Going beyond compliance
Operations interactions	Supervisors/leader
	Instructors/training
	Maintenance control
	Flight crew
	Cabin crew
	Dispatch
Formal safety indicators	Reporting system
	Response and feedback
	Safety personnel
Informal safety indicators	Accountability
	Technician's authority
	Professionalism

Source: Quoted from Gibbons and Thaden (2008)

3 Discussion

3.1 *Performance-based safety assessment*

The transition to a performance-based safety assessment scheme (ICAO, 2013) is deemed as a positive way to foster a culture of systems' effectiveness in addition to the required compliance to regulations; definitely, the fact that SMS components are documented by an organization does not guarantee that these are operated effectively. However, it seems that authorities have not clearly defined the different meanings between 'system effectiveness' and 'effective operation of a system'. The former regards the effects of the system on the organisation (i.e., safety outcome), whereas the latter refers to how much satisfactory a system is operated. It is of high interest that although SMS and occupational health and safety management systems were introduced in order to increase safety performance, there have been few studies to provide evidence for such a direct effect (e.g., Thomas, 2012). In addition, safety performance has been still measured widely by lagging indicators, discussed in the next sections, leading to the paradox of inferring effectiveness of present systems based on experience (e.g., accident rates).

The SMICG (2012) tool addresses effective operation of an SMS, whereby the aspects of presence (i.e., compliance to standards) and operability (i.e., evidence of ongoing activities) are fully comprehensive and easy to assess. However, when it comes to the terms of suitability and effectiveness, there is no guidance for the evaluation and these aspects remain vague. The basic Plan-Do-Check-Act quality cycle has been the foundation of the Eurocontrol's (2012) instrument that evaluates the effective operation of SMS, but includes statements that are difficult to be operationalised. For instance, requirements such as "The organisation has an effective mechanism in place to identify changes within the organisation that could affect regulatory processes..." and "Safety has a high priority during resource allocation..." lack guidance on how to assess effectiveness and prioritisation.

3.2 *Safety performance indicators*

ICAO's (2013) suggestion for the development of different indicators according to the administrative or operational activity level under consideration may be seen as an important guidance, since levels under evaluation must be comparable in their context, and common benchmarking basis is required. Regarding the approach of the Amey Vectra (2000), it contains highly subjective measurements of safety system fundamental components, indicated by scores such as 'partial', 'good', 'reasonable', etc.

Individual rate and frequency statistics for several activity types, operational units, etc. may comprise either benchmarking references or independent variables, which can be used to explore potential effects; each measurement may be performed as determined by the interests of the relevant organisational level. The development of extensive and countless indicators could reveal safety performance in detail and allow performance comparison over time and amongst several operational units. However, the following issues must be considered:

- Their establishment is a matter of a cost-benefit balance in terms of available technology (software and hardware), work force (availability and training), priorities and time allocation. Examples of a very detailed indicator in the aviation context, but

almost ineffective to calculate, might be the bird strikes per flock density over a flying area and per aircraft movements in the specific area; another example for a plant, might be the accidents due to chemical factors per population and per time unit exposed.

- Under the aforementioned concept, the suggestion of Bell et al. (2008) for recording the exposure to hazards, seems quite representative if the relevant records are kept. However, in cases of daily tasks not linked to a measured quantity, such a recording seems ambiguous in its practical implementation, since documentation of actual exposure (number of persons and time of exposure for each person) is almost unrealistic, even in small working areas.
- Regarding probability studies, it is important to comprehend that safety events are dependent on each other. Accidents and incidents take place in a constantly interactive environment, where each accident comprises a ‘dependency’ for the next accident. Precautionary and remedial measures resulting from a previous accident investigation have inevitably modified macro and/or micro conditions before the next accident occurs. Thus, neither discrete probability distributions based on figures of discrete exposure (e.g., trips, working days, flights, departures) nor continuous probability distributions referring to continuous exposure (working hours, flying hours, kilometres ran, etc.) can predict future safety events, because the fundamental principle of independency among safety events under concern would be violated.

The discussions amongst safety professionals regarding safety performance indicators (ASMSG, 2012) include ideas for indicators, which, however, do not always stand before criticism. Particularly:

- The average duration needed to close out issues derived from accident-incident reports, safety audits, safety meetings, hazard reports, etc. seems useful to get the big picture of a system’s reaction time; however, it misses the qualitative part of what type of issues are addressed faster and how the latter contributed in current safety problems. For instance, it is expected that accident reports’ communication to end-users, as a preventive measure, will demonstrate short duration; on the other hand, major and costly interventions leading to significant system reformations usually need considerable time to be applied. Therefore, such a measurement must refer to the same type of issues, such as procedural change in the operational and management levels, personal safety equipment, safety training, safety communication, etc.
- The number of safety related reports submitted through each organisation’s system, i.e., a mandatory or voluntary safety reporting system, is not always indicative of safety performance. What matters is the percentage of reports that allow either their follow up in order to attain more specific data on the reported problem (indicating also a trustful organisational culture), or their immediate treatment. Reports that cannot lead to remedial actions are, in fact, less useful and could not count in performance measurement. In addition, the reporting level must be considered; the value of an official report stems mainly from the rationale behind the reported hazard and the suggestion of remedial actions, which in combination offer a more detailed and holistic view of the problem. On the opposite side, an unstructured report from an end-user could be more realistic, providing a sense of the actual problem without

any filtering from the management. Therefore, the distance between such types of reporting levels, in terms of agreement, seems also essential for safety performance assessment.

- The value of counting safety meetings and recording their attendance seems low; most of the organisations periodically hold such meetings and pre-determine the participants. Rather few safety officers, managers, supervisors and other guests would avoid such meetings because of their reluctance to be characterised as ‘non-safety minded’ despite their actual commitment to safety. Hence, both meetings’ and attendants’ numbers are expected to be almost the same on an annual basis, hence, cannot indicate safety performance.
- Costs of safety events, either accidents or incidents, are an important indicator but may be seen as useful mainly for budget purposes. In fact, it is difficult to compare injuries and fatalities with material cost. One year the costs of accidents might had been €1.000.000 with three (3) major injuries and another year the costs might had summed up to €200.000 with one (1) injury that led to permanent disability; apparently, it is risky to compare safety performance between the aforementioned years from an accident cost viewpoint. Even though a ‘simulation’ tool could calculate the human losses in money terms, the psychological impacts on the persons directly involved and on the organisation cannot be exactly estimated and assigned a numerical value.
- The approach of weighing performance indicators in terms of their contribution to safety in order to provide managers with a ‘total safety performance indicator’ would be useful from a numerical aspect, but it is disputable how safety experts would decide the weight of each indicator and claim comparison with other organisations. Each enterprise could assign to indicators different priorities and prominence according to their its management culture; one director may think that the time to correspondent to safety issues is more important than the number of hazard reports submitted, or the opposite. Only a robust and commonly agreed scheme based on research results could seem useful in such a measurement scheme.

3.3 Safety audits and safety culture evaluation

The method of Fuller (1997) embodies the most important management aspects, and presents a reliable tool for assessing both overall performance and the influence of each management aspect on it. However, it was constructed in a strictly close form by using only numerical scale scores, lacks collection of qualitative data from employees, and refers only to high-level safety management activities (policy, organisation, planning, risk management, performance and auditing).

The tools discussed by the TC (2005), Gholami (2011), Burke et al. (2002), Downs (2003), Simon (2005), Gibbons and Thaden (2008), Kines et al.(2011), and Cameron and Duff (2007) use the results of audits, safety culture and climate assessments, etc. These tools might be accompanied with an instrument addressed to managers and safety professionals, who could score performance of specific safety actions and provide qualitative data to the assessors. On one hand, the scope of such a tool would be to complement auditors’ subjective views during site surveys, and on the other hand, assess

attitudes of the personnel responsible for planning, implementing, monitoring and reviewing of safety actions.

Safety culture is an important indicator for safety performance as discussed by Arezes and Miguel (2003), and Mitchell et al. (2002). However, taking into account that any changes in safety culture need time to be observed, safety culture measurement tools may comprise a reliable method for safety performance assessment only if they are used with a predetermined periodicity. Moreover, the factors that may have influenced the safety culture 'level' (e.g., management changes, working environment conditions, interrelationships, social expectations, rewards, and income) actually cannot be individually assessed, and potential disputes of what affected any measured culture change may arise. In addition, any safety culture indicators referring to periods between major internal changes and interventions regarding safety policies, procedures and practices, may mislead the organisations' decisions; such indicators are the result of continuous interaction of factors inside and outside the working environment and do not comprise the outcome of individual safety initiatives.

3.4 Cost-effectiveness measurements

The cost effectiveness of safety (benefits/costs) seems rather impossible to be objectively and realistically assessed as attempted by Gitelman et al. (2008), Moses and Savage (1997), and Marlow et al. (2004). More specifically:

- Apart from the easy to calculate material costs due to accidents, costs of fatalities cannot actually be quantified and mutually agreed in the organisational context. For example, the US Federal Aviation Administration and September 11 Commission set a range for such fatality costs from US\$3 to US\$8 million per individual on the scope of safety effectiveness assessment and insurance compensations to victims respectively (Schulman, 2006). However, it is rather arbitrary and ethically questionable for an organisation to document and publish such figures.
- Although insurance companies take into account victims' age, life expectancy income, marital status and dependents in order to calculate compensation costs, it is problematic to assess the psychological effects to the families of the victims and the physical effects of severe injuries across the organisation and the industrial sector in general.
- The quantification of moral and psychological factors is rather doubtful. Marlow et al. (2004) in order to compute the indirect costs suggested the subjective attribution of losses in productivity and quality to deviation from workers' 'ideal state'. Such arbitrary calculations cannot claim a solid basis and be widely accepted during internal and external benchmarking.
- The effects of safety policies, procedures, regulations and standards are continuous and multilevel. Their positive or negative impacts on the operational, professional, personal, social and national contexts cannot be objectively and robustly estimated.
- The cost benefit analysis regarding specific safety program components presumes that all resources are ideally managed and every established procedure is timely, qualitatively and completely accomplished. However, in real operations this is rather questionable since employees continuously strive to fulfil a variety of demands under

scarce resources. For instance, in the aviation context, foreign object damage (FOD) prevention program costs can be calculated by counting the working hours needed per day, the procurement and maintenance cost of magnetic and brush sweepers, the equipment's operational cost, etc. However, these costs may picture the effectiveness and efficiency of the FOD prevention policy if compared with the costs of FOD accidents, only if it is assumed that every mean was in service and all appointed staff worked ideally and at the same performance level every day, which is obviously unrealistic. The cost-benefit analysis becomes even more complex if the interdependencies of the various organisational activities are considered; for instance, FOD activities accomplishment by maintenance staff in a warm and heavy-duty day may undermine the quality of maintenance tasks following the FOD-related tasks.

3.5 *General issues*

The following issues, which were raised during the discussions held with safety professionals of the aviation organisation under reformation, are also important to be considered during development of safety performance measurements:

- Any measurement result might be variously interpreted either by different organisations or by departments and units within an organisation, depending on their culture, knowledge and experience in safety management. Example given, a decline of hazard report numbers might be perceived as either lack of problems or lack of confidence in safety authorities. Hence, it is of high importance for safety professionals to combine diverse views suitably and drive the formulation of meaningful and useful conclusions.
- The descriptive analysis of data (e.g., trends, frequency comparisons) must trigger surveys for validating observations, and be supported by factorial statistics in order to explore deeper systemic flaws, unusual trends and figures, extreme values, etc.
- In addition to the measurement of severity rates, indicators might include cause and factor rates appropriately adjusted to the exposure under concern (e.g., FOD/total flights, aircrew errors/total flying hours), depending on data availability. An illustration of the percentages of specific factors might be acceptable for use in sampling but would not be preferred whenever data for the overall organisation's activity is available.
- The rates of common factors and causes must be reported per accident type (e.g., flight, maintenance, ground handling) and in overall (e.g., supervision factor must be calculated both for all accidents and per accident type). This practice will allow depicting safety performance of each section and the organisation as a whole.
- Taking into consideration that staffing levels may fluctuate along time and across units, occupational accident rates may be more accurately calculated per person and per actual working hours. The rate of accident per 1,000 persons that is used by various organisations might be valid only to units with identical mission profiles and staffing levels while presuming an average level of job absence.

4 Conclusions

Most of the methods currently proposed for measuring safety performance are based on direct numerical data (accident rates, hazard and incidents reports, etc.), interview and questionnaire surveys that depict the safety status of an organisation (e.g., employees' satisfaction, behaviours, culture and practices), and cost-benefit analyses. Such approaches mainly provide information on 'what' the problem is (e.g., accident rates), and 'where' management could intervene (e.g., safety program elements that need consideration) in order to 'fix' safety problems. However, these methods do not explore and/or explain the 'why' and 'how' of the observed deficiencies (i.e., deeper safety performance assessment), in order to assist managers in addressing them. Hence, such assessment methods usually generate results that reflect the overall safety level and do not provide a detailed picture of the safety mechanisms and interdependencies that may better assist management in making more targeted and effective decisions. In addition, current tools that support a performance-based evaluation (e.g., SMICG, 2012; Eurocontrol, 2012) include vague measurement scales and cannot adequately assess effectiveness of an SMS beyond mere compliance.

Reactive measurements, such as indicators, correlations and benchmarking can be based on historical data such as annual reports, accident and incident investigation reports, safety meeting minutes, safety audit reports and risk registers. Although reactive, such safety performance indicators must be specific and understandable, measurable, achievable, relevant to the context under concern, and time bounded for their implementation and review. As Fung and Tam (1998) discussed such KPI must monitor the relevant critical success factors (CSF), the latter to be derived by a gap analysis between an ideal-complete SMS framework and the system in use.

5 Recommendations

Briefly articulating the discussion topics raised above, the following issues are viewed as critical to the establishment of an effective safety performance evaluation scheme:

- It is of high importance for authorities and organisations to define the substantial difference between system effectiveness and effective operation of a system. Under this need, new tools must be devised in order to link safety performance to effective operation of an SMS. Under this concept, the scope of SMICG (2012), Eurocontrol (2012) and any other similar tools must be amended in order to clarify that these refer to effective operation of SMS and not SMS effectiveness. Modern approaches to safety modelling, such as STAMP/STPA (Carroll et al., 2009) and FRAM (Hollnagel, 2012) can support the refinement of existing tools for evaluating effective operation of SMS, as proposed by de Boer and van de Maarel (2014).
- More research is required in order to provide evidence for the effects of SMS operation to organisational safety performance.
- Each organisation's monitoring authority needs to explicitly document its corresponding indicators, prioritise their assessment and allow flexibility in order to attain firstly the overall picture, and afterwards, the detailed one.

- The inevitable compromises appropriately adapted to the organisation's resources must be formulated for every indicator established (calculation of road accident rates with the hypothesis of comparable drivers' experience, road surface conditions, weather conditions, etc.).
- Safety climate assessments may be included in safety performance monitoring policy by establishing a comprehensive and valid qualitative and quantitative instrument based on the combination of already tested tools, such as those of Simon (2005), and Gibbons and Thaden (2008).
- Safety measurements between major organisational changes must take into account that every individual indicator embraces the continuous interaction of factors inside and outside the working environment and safety performance does not comprise the outcome of individual safety initiatives.
- Safety related cost-effectiveness calculations based on accident costs and savings due to safety measures seem highly subjective, and organisations are prompted to avoid their use as official performance indicators.
- The number of risk levels that were decreased and the ones that were increased during a specific period may be indicative of risk management effectiveness. New entries to the risk register and decreased risk levels might be seen as positive, and increased risk levels might count as negative to safety performance depending on their context.

References

- Adebiyi, K.A., Charles-Owaba, O.E. and Wahhed, M.A. (2007) 'Safety performance evaluation models: a review', *Disaster Prevention & Management*, Vol. 16, No. 2, pp.178–187.
- Al-Homoud, M.A. and Khan, M.M. (2004) 'Assessing safety measures in residential buildings in Saudi Arabia', *Building Research & Information*, Vol. 32, No. 4, pp.300–305.
- Amey Vectra, L. (2000) *Development of a Health & Safety Performance Measurement Tool Contract Research Report 309/2000*, Health & Safety Executive, Warrington.
- Arezes, P.M. and Miguel, A.S. (2003) 'The role of safety culture in safety performance measurement', *Measuring Business Excellence*, Vol. 7, No. 4, pp.20–28.
- Aviation Safety Management Systems Group (ASMSG) (2012) *Safety Performance Indicators*, 16 December [online] <http://www.linkedin.com> (accessed 10-12-2012).
- Barrie, D. (1990) 'Performance indicators', *Work Study*, Vol. 39, No. 5, pp.22–26.
- Bell, K.L., Nigel, R., O'Connell, M.S. and Reeder, M. (2008) 'Predicting & improving safety performance', *Industrial Management*, March/April, pp.12–16.
- Bititci, U., Dorfler, V., Garengo, P. and Nudurupati, S. (2012) 'Performance measurement: challenges for tomorrow', *International Journal of Management Reviews*, Vol. 14, No. 3, pp.305–327.
- Boyle, A.J. (2008) 'The collection & use of accident & incident data', in Channing, J. and Ridley, J. (Eds.): *Safety at Work*, Butterworth-Heinemann, UK.
- Boyle, B., Broadribb, M.P. and Tanzi, S.J. (2010) 'Safety performance indicators: cheddar or Swiss? How strong are your barriers? (One's company's experience with process safety metrics)', *Loss Prevention Bulletin*, pp.29–40.
- Burke, M.J., Sarpy, S.A., Smith-Crow, K. and Tesluk, P.E. (2002) 'General safety performance: a test of a grounded theoretical model', *Personnel Psychology*, Vol. 55, No. 2, pp.429–457.

- Cameron, I. and Duff, R. (2007) 'Use of performance measurement & goal setting to improve constructions managers' focus on health & safety', *Construction Management & Economics*, Vol. 25, No. 8, pp.869–881.
- Carroll, J., Dulac, N., Leveson, N. and Marais, K. (2009) 'Moving beyond normal accidents and high reliability organizations: a systems approach to safety in complex systems', *Organization Studies*, Vol. 30, Nos. 2–3, pp.227–249.
- Channing, J. and Ridley, J. (2008) *Safety at Work*, 7th ed., Butterworth-Heinemann, UK.
- Davies, J., Ross, A., Wallace, B. and Wright, L. (2003) *Safety Management: A Qualitative Systems Approach*, Taylor & Francis, UK.
- de Boer, R.J. and van de Maarel, R. (2014) 'Applying STAMP to improve evaluation of SMS', *2nd European STAMP Workshop*, pp.607–611.
- Downs, D. (2003) 'Performance improvement: safety & environmental management system assessment', *Professional Safety*, November, pp.31–38.
- Easter, K., Hegney, R. and Taylor, G. (2004) *Enhancing Occupational Safety & Health*, Butterworth-Heinemann, UK.
- Eurocontrol (2012) *Effectiveness of Safety Management*, Brussels.
- European Aviation Safety Agency (EASA) (2014) *A Harmonised European Approach to a Performance Based Environment*, Cologne.
- Ferrett, E. and Hughes, P. (2007) *Introduction to Health & Safety in Construction*, 2nd ed., Butterworth-Heinemann, UK.
- Ferrett, E. and Hughes, P. (2011) *Introduction to Health & Safety at Work*, 5th ed., Butterworth-Heinemann, UK.
- Fuller, C. (1999) 'Benchmarking health & safety performance through company safety competitions', *International Journal for Benchmarking*, Vol. 6, No. 5, pp.325–337.
- Fuller, C.W. (1997) 'Key performance indicators for benchmarking health & safety management in intra- & inter-company comparisons', *Benchmarking for Quality Management & Technology*, Vol. 4, No. 3, pp.165–174.
- Fung, I.W. and Tam, C. (1998) 'Effectiveness of safety management strategies on safety performance in Hong Kong', *Construction Management & Economics*, Vol. 16, No. 1, pp.49–55.
- Gholami, S. (2011) 'Total safety performance evaluation management', *Interdisciplinary Journal of Contemporary Research in Business*, Vol. 3, No. 2, pp.1185–1197.
- Gibbons, A.M. and Thaden, T.L. (2008) *The Safety Culture Indicator Scale Measurement System*, University of Illinois – Institute of Aviation – Human Factors Division, Illinois.
- Gitelman, V., Hakkert, A.S., Winkelbauer, M., Papadimitriou, E. and Yannis, G. (2008) 'Testing a framework for the efficiency assessment of road safety measures', *Transport Reviews*, Vol. 28, No. 3, pp.281–301.
- Goglia, J., Halford, C.D. and Stolzer, A.J. (2008) *Safety Management Systems in Aviation*, Ashgate, UK.
- Hollnagel, E. (2012) *FRAM – The Functional Resonance Analysis Method*, Ashgate, UK.
- International Civil Aviation Organization (ICAO) (2013) *Safety Management Manual*, Doc. 9859, Canada.
- Kaufman, R. (2011) 'Toward a generic process for individual and organizational performance improvement and contribution', *Performance Improvement*, Vol. 50, No. 9, pp.32–40.
- Kines, P., Lappalainen, J., Mikkelsen, K.L., Olsen, E., Pousette, A., Tharalkdsen, J. and Torner, M. (2011) 'Nordic safety climate questionnaire (NOSACQ-50): a new tool for diagnosing occupational safety climate', *International Journal of Industrial Economics*, Vol. 41, No. 6, pp.634–646.
- Marlow, P., Oxenburgh, A. and Oxenburgh, M. (2004) *Increasing Productivity & Profit through Health & Safety*, Taylor & Francis, UK.

- Mitchell, A., Sharma, G., Thaden, T., Wiegmann, D.A. and Zhang, H. (2002) *A Synthesis of Safety Culture & Safety Climate Research*, University of Illinois – Aviation Research Lab, Illinois.
- Moses, L.N. and Savage, I. (1997) 'A cost-benefit analysis of US Motor Carrier safety programs', *Journal of Transport Economics & Policy*, Vol. 31, No. 1, pp.51–67.
- Parker, C. (2000) 'Performance measurement', *Work Study*, Vol. 49, No. 2, pp.63–66.
- Safety Management International Collaboration Group (SMICG) (2012) *Safety Management System Evaluation Tool*.
- Sagar, M., Sagar, S. and Yadav, N. (2013) 'Performance measurement and management frameworks: research trends in the last two decades', *Business Process Management Journal*, Vol. 19, No. 6, pp.947–970.
- Schulman, A.B. (2006) 'Financial stability & airline safety: relationships, causes & consequences', *International Journal of Applied Aviation Studies*, Vol. 6, No. 2, pp.249–270.
- Simon, S.I. (2005) *Safety Culture Assessment as Transformative Process*, Culture Change Consultants, NY.
- Stapenhurst, T. (2009) *The Benchmarking Book: A How-to-Guide to best Practice for Managers & Practitioners*, Elsevier, UK.
- Stranks, J. (1994) *Management Systems for Safety*, Pearson Education, UK.
- Stranks, J. (2006) *Health & Safety Handbook*, Kogan Page Ltd., UK.
- Stranks, J. (2008) *Health & Safety at Work: An Essential Guide for Managers*, Kogan Page Ltd., UK.
- Thomas, M. (2012) *A Systematic Review of the Effectiveness of Safety Management Systems*, Australian Transport Safety Bureau, Australia.
- Transport Canada (TC) (2002) *Safety Management Systems for Flight Operations & Aircraft Maintenance Organizations*, Transport Canada, Ottawa.
- Transport Canada (TC) (2004) *Safety Management Systems for Small Aviation Operations*, Transport Canada, Ottawa.
- Transport Canada (TC) (2005) *Safety Management System Assessment Guide*, TP 14326E, Transport Canada, Canada.