

# Exploring the Diversity in Safety Measurement Practices: Empirical Results from Aviation

Steffen Kaspers\*, Nektarios Karanikas, Alfred Roelen, Selma Piric, Robbert van Aalst &  
Robert J. De Boer<sup>1</sup>

Amsterdam University of Applied Sciences

Weesperzijde 190, 1097 DZ Amsterdam

\*Tel: +31621155891

\*Email: s.e.kaspers@hva.nl

doi:10.5296/jss.v2i2.10437

URL: <http://dx.doi.org/10.5296/jss.v2i2.10437>

## **Abstract**

A literature review, which was conducted during the research project “Measuring Safety in Aviation – Developing Metrics for Safety Management Systems”, identified several problems and challenges regarding safety performance metrics in aviation. The findings from this review were used to create a framework for interviewing 13 companies in order to explore how safety performance is measured in the industry. The results from the surveys showed a wide variety of approaches for assessing the level of safety. The companies encounter and/or recognise problematic areas in practice when implementing their safety management. The findings from the literature review are partially confirmed and it seems that the current ways of measuring safety performance are not as straight forward as it might be assumed. Further research is recommended to explore alternative methods for measuring aviation safety performance.

**Keywords:** Safety performance, Safety metrics, Safety Management System

## 1. Introduction

A literature review, which was performed as part of the research project “Measuring Safety in Aviation – Developing Metrics for Safety Management Systems” (Aviation Academy, 2014), identified views and practices on safety metrics (Kaspers et al, 2016). This review concluded to the following:

1. Safety is widely seen as avoidance of failures and is managed through the typical risk management cycle which includes the stages of hazard identification, risk assessment, risk mitigation and risk monitoring. Under this concept:
  - a. Hazards are identified through a spectrum of sources.
  - b. Risk assessment is predominately based on probabilistic approaches, which employ estimations of likelihood and severity.
  - c. Risk mitigation or elimination is achieved through barriers of various types.
  - d. Risks are actually monitored through the same sources that hazards are identified.
2. Safety metrics can be split in two groups: safety process and outcome metrics.
  - a. Safety process metrics are linked with operational, organizational and Safety Management System (SMS) activities.
  - b. Outcomes are occurrences of any severity category (i.e. accident, serious incident, incident). These outcomes are used by the industry to develop respective indicators (e.g., number of occurrences per departure) for measuring safety performance. However, thresholds for incidents and serious incidents are not clearly defined. In addition, accidents and incidents are infrequent considering the amount of activities, therefore they cannot be seen as a useful indication of current safety level.
3. There is a lack of standardization for the development of safety metrics and there is no explicit reference to quality criteria regarding their design. Standards have mandated the transition from compliance to performance based evaluations of safety, but this concept is not yet backed with tools and techniques
4. Safety culture is seen as either an outcome indicator or process indicator.
5. There is limited empirical evidence about the relationship between process and outcome metrics and the link between those often relies on plausible reasoning. Such reasoning is principally based on linear safety/accident models. Systemic models have been introduced in literature but they have not been extensively applied to the industry.

The literature review results were used to create a framework for interviewing 13 organisations as a means to explore the practices of safety performance measurement across the research project partners. This paper presents the findings of the respective surveys.

## 2. Methodology

In order to answer the overarching question “To what extent are the results from the literature review evident in industry practice?”, sub-questions were formulated as shown in Table 1.

Table 1. Research Sub-questions

No	Sub-question	Correspondence with literature review findings
Q1	How do the companies perform risk assessments?	1a, 1b
Q2	What types of safety metrics do companies use and are those metrics comparable?	2
Q3	Do the safety metrics used by the companies adhere to the quality criteria mentioned in the literature?	3
Q4	How is safety culture seen in a SMS?	4
Q5	What are the safety paradigms/views used in practice?	5

The research team interviewed safety managers and professionals from 13 European aviation companies represented by one to three safety staff. Out of the 13 companies, seven were large (i.e. >250 employees) and six companies fell under the category of Small and Medium-sized Enterprises (SME) (i.e. < 250 people). The participating companies are distributed across four domains: Flights Operators (Flight Ops, N=7), Air Navigation Service Providers (ANSP, N=2), Ground Service Provider (GSP, N=1) and Maintenance, Repair and Overhaul service providers (MRO, N=3). The interviews were conducted between February and April 2016 and lasted 4 to 6 hours. The interview team consisted of two research team members and one graduate student, but in two cases the interview was conducted by only one researcher due to practical limitations. Ten companies allowed audio recording of the interviews.

Each interview included four parts:

1. A presentation of the results from the literature review (Kaspers et al, 2016) by the research team.
2. An explanation by the company on how they implement their SMS.
3. A first interview focusing on what, how and why things are measured in regard to safety.
4. A second interview to explore what SMS related data companies record but might not directly use in their safety metrics. This part of the interview was structured according to the SMS elements described in the Safety Management Manual (ICAO, 2013).

The interview notes were cross-checked by all members of the interview team and when inconsistencies were indicated, the audio files were consulted. The notes were subject to a template analysis based on the findings of the literature review (Kaspers et al, 2016) and the correspondence presented in Table 1.

### 3. Results from Qualitative Data Analysis

#### 3.1 Risk Assessment and Safety Metrics

The inputs used by the companies for their risk assessment are shown in Table 2. Those inputs constitute also the basis for measuring safety; the left column of the table refers to the measurements each company uses.

Table 2. Inputs to risk assessment

	Company Size		Activity Domain			
	Large (N=7)	SME (N=6)	Flight Ops (N=7)	ANSP (N=2)	GSP (N=1)	MRO (N=3)
Compliance monitoring	7	6	7	2	1	3
Operational Data (Flight Data Monitoring & Air Navigation Service Provider Data Monitoring)	5	1	4	2		
Line Operations Safety Audits	2	1	2		1	
SMS Maturity score	2			2		
Feedback from training	1		1			
Voluntary reporting	7	6	7	2	1	3
Safety outcomes	7	6	7	2	1	3
Trends of hazards, etc. over time	7	6	7	2	1	3

The results in Table 2 along with the context provided by the companies during the interviews showed that:

- All companies use compliance monitoring based on internal and/or external audits. However, one company mentioned that “...during an audit everybody puts on their best show, and after the inspectors leave, everybody goes back to normal work”.
- Large companies mainly use operational data for their risk assessment. SMEs do not always have technical capabilities to provide this type of data and according to the Commission Regulation (EU) No 965/2012 Annex III (EC, 2012) flight data monitoring is not required for small aircraft.
- Three out of the 13 companies use Line Operations Safety Audits (LOSA) as input to their risk assessment. The concept of LOSA is that trained observers evaluate staff during normal activities. These observers provide feedback to the employees and the organization as a means to continuously improve safety. LOSA is an internal means of compliance and detection of deviations along with their context.
- Two ANSPs assess their SMS with a self-scoring maturity score (Eurocontrol 2009).
- One company uses feedback from safety training as an input for the SMS process.
- All companies have a system in place where employees can report any safety related case. The interview results indicated that such a formal reporting system is not consistently used in small companies, and coffee table talks among employees comprise a basic source of information. However, for large companies reporting is seen as a valuable resource for SMS improvement. Results from the reporting system are used for three purposes: identification of hazards, contextualization of occurrences, and indication of safety culture.

- All companies monitor their safety outcomes. However, the participants mentioned that the lack of clarity and specific thresholds in aviation standards and regulations can result in different interpretations across and within companies.
- All companies look for trends in their data over time and their monitoring intervals differ. None of the companies reported the establishment of predetermined alert limits and trends are evaluated qualitatively; if a trend is recognised, the company might act or not.

After data are collected, most companies assess the risk level with a risk matrix. Estimates of probabilities and severities are based on past cases or expert judgment. The resulting risk level determines priorities, which might be reprioritised depending on additional context. The interviews showed that:

- Nine companies use a 5x5 matrix, whereas the two ANSPs use their own 6x5 matrix with an additional row/column for undefined/non-assessed risks. Two out of the three MRO companies did not explicitly state the use of such a matrix.
- One air operator stated that the risk assessment is completely arbitrary, because the results are highly dependent on the expert who is available to assess the risk(s).
- One SME felt unsure about the use of its risk matrix and is interested in a more objective manner to assess risks and comparisons with companies.

### 3.2 Criteria for safety metrics development

The criteria companies employ for developing their safety metrics are presented in Table 3.

Table 3. Methods for creating safety metrics

	Company Size		Activity Domain			
	Large (N=7)	SME (N=6)	Flight Ops (N=7)	ANSP (N=2)	GSP (N=1)	MRO (N=3)
Measure what is measurable	3		1		1	1
Based on expert judgement, standards, and professional knowledge	3	4	4	2		1
Trial and error	1			1		
Indicators change over time	1			1		
SMART	2			1	1	

One small MRO mentioned that it assesses safety in a qualitative manner without using formal metrics. According to the findings for the rest of the companies:

- Companies follow the guidance of standards (e.g. ICAO, 2013), own professional knowledge and/or the practices shared in the industry.
- Three large companies “measure everything that can be measured”.

- One company uses metrics based on a trial and error approach. They look for metrics that are relevant to the process of concern; if the metrics seem suitable, they are maintained and tracked, otherwise they are replaced. However, criteria for suitability were not stated.
- One company mentioned that their metrics change over time.
- Three companies mentioned the use of SMART criteria.

An evaluation of the Table 3 metrics against the quality criteria found in literature (Kaspers et al, 2016) showed that:

- There is no explicit theoretical framework supporting the metrics.
- Most of the metrics are specific and measurable but those characteristics depend on the instrument used for the data collection and the interpretation of the data analysis results.
- Validity of the metrics is only partially met due to factors such as lack of a systemic approach, subjective implementation of the respective tools and ambiguous definitions.
- No metric is completely immune to manipulation.
- The practicality and cost-effectiveness of the metrics is dependable on the amount and nature of data collected and analysed in relation with the available resources.
- The reliability of the metrics is not guaranteed due to subjective evaluations the most of the metrics require.
- The frequency of monitoring is the main factor influencing the sensitivity to changes of conditions.

### 3.3 Safety culture and models

Nine companies mentioned the importance of culture by referring to one or more types of culture, such as just culture, safety culture or reporting culture (Table 4). However, none of the companies measure their culture consistently; only one ANSP occasionally assessed the safety culture, but this was not viewed as a regular safety metric by the specific company.

Table 4. Culture types mentioned by the companies

	Size		Domain			
	Large (N=7)	SME (N=6)	Flight Ops (N=7)	ANSP (N=2)	GSP (N=1)	MRO (N=3)
Culture (all)	6	3	5	1	1	2
Safety culture	5	2	3	1	1	2
Reporting culture	1	1	2			
Just Culture	2	1	2	1		

As shown in Table 5, the companies mainly consider safety in a linear, direct cause-effect way. Only three large companies use both systemic and linear models to analyse incident and accidents, but the choice of the model depends on the resources available.

Table 5. Models mentioned by the companies.

	Size		Domain			
	Large (N=7)	SME (N=6)	Flight Ops (N=7)	ANSP (N=2)	GSP (N=1)	MRO (N=3)
Systemic models	3		1	2		
Linear models	6	3	4	2	1	3

#### 4. Discussion

The results are discussed in correspondence with the questions listed in Table 1 and in accordance with the contextualized information the companies offered during the surveys.

##### 4.1 How do the companies perform their risk management?

Each company implements SMS differently and develops the respective processes according to their operational profile. All companies that are obliged to implement a SMS follow the risk cycle included in the Safety Management Manual (SMM) (ICAO, 2013) and they use risk matrices. However, some companies recognised that the risk assessment method is not adequately objective. In the lack of reliable historical data, the estimation of probability and severity of an occurrence is initially performed by a person. Although guidance to limit the effect of biases exists (e.g., Cooke, and Goossens, 2000), the existence of biases was acknowledged by several companies and literature (Duijm, 2015; Hubbard et al., 2010; Karanikas & Kaspers, 2016).

Also, SMEs acknowledged a lack of confidence in the risk area limits they have set in their matrices since uniformity and standardization is missing. Therefore, while standards allow companies to tailor risk matrices to their operations, little guidance is provided for developing and using such matrices. This potentially leads to a wide variety of matrix areas and risk measurements, accompanied by their own definitions, this not allowing a benchmarking amongst companies.

##### 4.2 What types of safety metrics do companies use and are those metrics comparable?

Companies use both safety process metrics and outcome metrics in the frame of their safety management. Process data are used to improve safety outcomes but they are not exploited to assess whether SMS processes in general perform adequately. Hence, companies use their safety metrics as sources for identifying hazards that are subject to risk management (ICAO, 2013).

All companies collect data about compliance, reporting, outcomes and trends. The results from the survey suggest that:

- Reporting seems to be more formalised at large companies, this possibly attributed to the need to streamline the dataflow. For SMEs, it seems easier to share such information informally; stories are shared around a coffee table. Nonetheless, regardless of size,

reporting is highly dependent on perceptions about what is worthwhile to be shared.

- SMEs have limited access to operational data due to constraints of available aircraft technology and company resources for analysis.
- Large companies look for trends more systemically, at more regular and smaller intervals compared to SMEs. This can be attributed to differences in resources, volume of operations and staffing of safety departments.
- Compared to large companies, SMEs experience fewer safety events. Large companies have generally more data about outcomes in terms of raw numbers. Both large and small companies do not consistently connect and maintain SMS data for their use in safety metrics
- A relation between safety processes and outcomes is assumed, and both types of metrics are compared with past figures. Companies seek for improvements, but they have not established control limits for safety metrics, even though this is required according to the standards (ICAO, 2013).

Furthermore, safety metrics can be used both proactively and reactively. Voluntary reports are used on a case-by-case basis for investigating the occurrences and derive lessons for the future (i.e. a reactive approach). Only one company stated proactive use of voluntary reporting as a means to identify concerns of employees and to assess whether personnel actively participate in a SMS.

Also, safety metrics used by companies do not allow valid comparisons across different organisations because such metrics depend on the data collected by each company and are not based on a common standard. Even more importantly, the widely-used ratios of safety events, and especially the ones of medium and low severity, cannot be directly compared across and within companies due to different interpretations of severity thresholds. For example, a taxiway take-off could be classified differently by company analysts depending on the variables they consider (e.g., other traffic, visibility conditions). Also, the context can affect the views of the air operator, the flight crew and the ANSP, all of those possibly classifying that event differently based on their own “process/subsystem”.

#### *4.3 Do the safety metrics used by the companies adhere to the quality criteria mentioned in the literature?*

In general, safety metrics used by the companies are not grounded on the whole set of the quality criteria suggested in the literature (Kaspers et al, 2016). Instead, participants follow a pragmatic approach to the development of safety indicators and these mainly stem from practice and expert judgment; as soon as metrics seem meaningful to a company, they are maintained and monitored. Thus, without predetermined criteria, service providers judge the quality of their current metrics based on expectations and common practice. This finding might reflect that companies focus on realising their objectives rather than examining the rigorousness of their metrics.

More specifically, several companies mentioned the SMART criteria (Doran, 1981) but

validity, cost-effectiveness and the existence of a theoretical framework were not mentioned as criteria used for developing safety metrics. Also, the criterion for sensitivity to changes in conditions cannot be ensured in the count of safety outcomes since the latter regard specific findings and events that are not completely repeatable under the dynamic nature of operations. Especially regarding outcome metrics, the ambiguous definitions do not allow uniformity when classifying events and the widely-used event rates are not directly comparable. Even within some of the companies it is sometimes hard to reach consensus on classifying a certain occurrence.

Therefore, few quality criteria are partially or fully met and some of those metrics depend on the company resources and measurement instruments. Some explanatory and summative remarks on the metrics used by the participants are the following:

- Compliance is based on the concept that adherence to the rules ensures a minimum level of safety, but half of the companies stated that safety is more than just compliance. There were connotations that simply following the rules does not guarantee safety. First, rules can be realised through various means, the acceptance of the latter being subject to the knowledge and skills of the auditor. Second, rules do not apply to every situation, since conditions and/or the context are forevermore changing. Third, there might be situations where rules are contradictory and decisions about balancing competing goals rely on the company and/or the end-user.
- Operational data monitoring might be useful to assess frequencies of events but raw data do not capture the context, which can be provided by reports. Thus, in the frame of effective safety management, numbers and coding of events must trigger exploration of the conditions; this depends on available resources.
- The effectiveness of LOSA depends on the instrument used, the skills of the observer and the perceptions and adaptive behaviour of the subjects.
- The SMS maturity score used by ANSPs is a quite abstract and subjective metric since it is self-scoring.
- Reporting that provides context to occurrences is seen as important and can reveal new hazards. However, the value of reporting as safety metric is debatable. An increased number of reports might indicate a better reporting culture and/or that more occurrences happened. Also, the quality of the reports determines the opportunities for learning; if only basic information is given, this is just entered in a database and used in statistics. If a report is rich in terms of context, data, views and decisions made, much may be learned. Furthermore, if companies demand a certain amount of reports from their employees, this might be seen as a requirement for compliance with regulatory requirements that dictate the operation of a 'voluntary' reporting process.

#### *4.4 How is safety culture seen in a SMS?*

Although the companies mention culture as an important element for determining the level of safety, none of the companies measures culture with a predetermined periodicity. The level of safety culture was indirectly indicated through the participation and response of staff to SMS

initiatives. For example, safety culture might be indicated through a comparison of FDM triggers with the amount of corresponding voluntary reports. Sometimes a safety manager's own perception about the willingness of employees to talk openly signalled a mature safety culture to the companies. Although this can provide some indication, it can be subject to biases.

#### *4.5 What are the safety paradigms/views used in practice?*

The metrics that are used by the companies suggest a focus on negative outcomes, or situations that deviate from normal operations. This indicates that industry practice is based on traditional views on safety. This is an expected result since the guidance material from ICAO (2013) refers to linear safety concepts such as the Swiss Cheese model (Reason, 1990). However, there is recognition by the companies that the current metrics do not suffice and that compliance alone is not safety. Only three companies mentioned the use of systemic models for assessing safety performance. The little consideration of newer safety/accident models might be attributed to the lack of analytical tools or their complexity.

### **5. Conclusions**

The results of the analysis of qualitative data partially verified the findings from the literature review (Kaspers et al, 2016). It was confirmed that:

- Safety is managed through the risk management cycle described in standards, and companies acknowledge the limitations of current risk assessment techniques.
- The safety data collected by the companies retrofit the risk management processes.
- Safety outcomes are used for safety performance, but their definitions are ambiguous.
- Accidents and incidents are infrequent events and cannot constitute reliable measurements of safety performance.
- Companies do not use predefined quality criteria for the design of their safety metrics; each company uses metrics that are specifically tailored to its organisation.
- Traditional approaches are used for safety management, and most of the companies follow linear safety concepts, which have not probably replaced with systemic ones due to the confidence and trust to industry practice and experience for years. Few companies explore approaches based on systemic models, but it seems that a smooth transition to such models is needed in order to render them doable and realistic.
- Safety culture is seen as important part of safety management.

Contradictory to the expectations raised from the literature review, the research revealed that:

- Current safety metrics are not grounded on sound theoretical frameworks and, in general, do not fulfil the quality criteria as proposed in literature.
- Safety culture is not a consistent part of safety metrics and, therefore, not assessed.
- The companies collect data related to their SMS processes, but such data are not associated with SMS metrics; some of the processes are performed but not measured.
- The data used differ across companies depending on perceptions, safety models used implicitly or explicitly, and available resources.

- SMS assessment is based on a compliance-based approach.

Overall, the findings of this study indicate a wide variety of approaches to safety performance and several problematic areas in aviation, which can be also investigated into other industry sectors by following a similar methodology. Nonetheless, further research is needed to strengthen or disprove the empirical evidence of the current study. As part of the next project phases we are going to research alternative manners to measure safety performance by exploring several options, including the application of systemic models.

### **Acknowledgments**

The research is co-funded by the Nationaal Regieorgaan Praktijkgericht Onderzoek SIA (SIA, 2015). The research team would like to express their deep thanks to:

- The companies which participated in the surveys: Helicentre, JetSupport, KLM (Royal Dutch Airlines), KLM Cityhopper, Life Line Aviation, LNVL (Air Traffic Control the Netherlands), MUAC (Maastricht Upper Area Control Centre), Olympus Airways, SAMCO, Sky Service Netherlands, Transavia.
- The Aviation Academy graduates: Hayo Brugmans, Marijn van der Deure, Mitchel Nauman and Shivaïd Ishaak.
- The members of the knowledge experts group of the project: CAA NZ: Charlotte Mills, EASA: John Franklin, Kindunos: John Stoop, KLM Cityhopper: Ewout Hiltermann, Klu / MLA: Ruud Van Maurik, Purdue University: Julius Keller, Team HF: Gesine Hofinger, TU Delft: Alexei Sharpanskykh.

### **References**

Aviation Academy. (2014) "Project Plan RAAK PRO: Measuring safety in aviation – developing metrics for Safety Management Systems", Hogeschool van Amsterdam, Aviation Academy, The Netherlands.

Cooke, R.M., Goossens, L.H.J. (2000). Procedures guide for structured expert judgement, EURATOM document EUR 18820EN, European Communities.

Doran, G. T. (1981). "There's a S.M.A.R.T. way to write management's goals and objectives". Management Review. AMA FORUM. 70 (11): 35–36.

Duijm, N.J., (2015). Recommendations on the use and design of risk matrices. Safety Science, 76, 21 - 31

EC, (2012), Commission Regulation (EU) No 965/2012 of 5 October 2012 laying down technical requirements and administrative procedures related to air operations pursuant to Regulation (EC) No 216/2008 of the European Parliament and of the Council (Official Journal of the European Union). Retrieved from <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012R0965&from=EN>.

EC, (2014). Regulation (EU) No 376/2014 of The European Parliament and of the Council (Official Journal of the European Union). Retrieved from [http://eur - lex.europa.eu/legal -](http://eur-lex.europa.eu/legal-)

content/EN/TXT/PDF/?uri=CELEX:32014R0376&from=EN

Eurocontrol. (2009) ATM safety framework maturity survey for ANSPs, 3rd SAFREP TF report to Provisional Counsel Appendix 1 2009.

Hubbard, D., & Evans, D., (2010). Problems with scoring methods and ordinal scales in risk assessment. *Ibm Journal of Research and Development*, 54, 3, 2:1 - 2:10.

ICAO (2013). Doc 9859, Safety Management Manual (SMM) (3rd Ed.) International Civil Aviation Organization. Montréal, Canada.

Karanikas, N & Kaspers, S. (2016). Do Experts Agree When Assessing Risks? An Empirical Study. Proceedings of the 50th ESReDA Seminar, 18-19 May 2016, Seville, Spain.

Kaspers, S., Karanikas, N., Roelen, A.L.C., Piric, S., & de Boer, R. J. (2016). Review of Existing Aviation Safety Metrics, RAAK PRO Project: Measuring Safety in Aviation, Project S10931, Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands.

Reason, J. (1990). *Human error*. New York: Cambridge University Press.

SIA. (2015). Besluit inzake aanvraag subsidie regeling RAAK-PRO 2014 voor het project Measuring Safety in Aviation – Developing Metrics for Safety Management Systems ' (projectnummer:2014-01-11ePRO). Kenmerk: 2015-456, Nationaal Regieorgaan Praktijkgericht Onderzoek SIA. The Netherlands.

### **Copyright Disclaimer**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).