The AVAC-SMS Metric for the Self-assessment of Maturity of Aviation Safety Management Systems

Nektarios Karanikas[1,*], Selma Piric[1], Robert Jan de Boer[1], Alfred Roelen[1,2], Steffen Kaspers[1] and Robbert van Aalst[1]

[1]Aviation Academy, Faculty of Technology, Amsterdam University of Applied Sciences, the Netherlands
[2]Netherlands Aerospace Center, the Netherlands

Abstract

This paper introduces the AVAC-SMS maturity metric and its accompanying tool which were developed in the frame of a research project with the aim to suggest new safety metrics, especially for Small-Medium Enterprises (SMEs). The metric is based on the ICAO Safety Management Manual, it was designed by applying the Systems-Theoretic Process Analysis (STPA) technique and it was reviewed by companies, authorities and field experts. It can be used to assess the institutionalisation, capability and effectiveness of an aviation SMS by following a systematic approach that employs the use of information from the safety department, managers and employees of an organisation. The AVAC-SMS maturity metric is uniform for the aviation sector, customisable to the size and complexity of the organisation, and results in numerical scores that can be used to monitor SMS maturity levels over time or perform benchmarking among companies.

Key Words: SMS Assessment; SMS Evaluation; SMS Maturity; Safety Management

1 Introduction

In academic literature (e.g., Dekker, 2011; Leveson, 2015) and industry standards (e.g., ICAO, 2013; FAA, 2006) it is argued that safety is affected by

*Corresponding author: n.karanikas@hva.nl, +31621156287

the gaps between Work-as-Imagined (WaI), which reflects how tasks must be accomplished (e.g., regulations, standards, procedures) and Work-as-Done (WaD) that represents how work is actually performed on the work floor. On one hand, it is recognised that such gaps are inevitable, but on the other hand, when they are uncontrolled and grow continuously, the system might drift to undesired performance levels. Although safety initiatives such as Safety Management System (SMS) audits aim at capturing the WaI-WaD gaps, to the knowledge of the authors there have been no relevant metrics to sufficiently and evidently, assess and quantify system behaviour and performance.

In September 2015, the Aviation Academy of the Amsterdam University of Applied Sciences initiated a research project entitled "Measuring Safety in Aviation – Developing Metrics for Safety Management Systems". The research aims to identify ways to measure safety proactively in scientifically rigorous, meaningful and practical ways without the benefit of large amounts of data (Aviation Academy, 2014). The primary focus of the researchers was to derive metrics that correspond to different instances of the distance between WaI and WaD with the premise that if this distance is large, changes must be induced to both or either of them. During the first phase of the project, the research, amongst others, concluded that SMS assessment is yet based on compliance-based approaches (Kaspers et al., 2016b, 2016c).

An SMS aims at assuring the safety of operations through effective management of safety risks. The four components of SMS and their elements are (ICAO, 2013):

- Safety Policy & Objectives, including the elements of Management Commitment and Responsibility, Safety Accountabilities, Appointment of Key Safety Personnel, Coordination of Emergency Response Planning, and SMS Documentation.
- Safety Risk Management that consists of the elements Hazard Identification, and Safety Risk Assessment and Mitigation.
- Safety Assurance whose elements are Safety Performance Monitoring and Measurement, The Management of Change, and Continuous Improvement of the SMS.
- Safety Promotion which incorporates the elements of Training and Education, and Safety Communication.

The industry has recognized the need to move from a compliance-driven assessment of SMS to a performance-based evaluation scheme (ICAO, 2013; EASA, 2014). Tools such as the Safety Management System Evaluation Tool

developed by the Safety Management International Collaboration Group (SMICG, 2012), and the Effectiveness of Safety Management (EoSM) instrument, which was devised by the Eurocontrol (2012), have been introduced to support the transition from compliance-driven to performance-based assessment but they include vague measurement scales and do not address the connections and dependencies of SMS processes (Karanikas, 2016). The recently launched Management System Assessment Tool (EASA, 2017) is based on the SMICG approach and Annex 19 of the International Civil Aviation Organization (ICAO, 2016). The particular tool is designed to be used for checking whether SMS parts are present (i.e. designed/documented), suitable for the size and complexity of the organisation, operational (i.e. implemented) and effective (i.e. achievement of desired outcome and impact on operations). A first examination of the EASA tool indicates that each assessment topic includes several sub-topics (i.e. multi-barreled questions), there is no guidance for the assessment of SMS suitability, in various cases the operationalisation and effectiveness of SMS parts are assessed in a similar manner, and, in general, effectiveness seems examined in a non-systematic way through indicative questions that have been derived from past knowledge and experience.

Therefore, although current SMS tools used by the aviation industry adhere to ICAO standards, introduce the transition from merely checking the existence of SMS elements and processes to considering the sufficiency of their output and indicating necessary improvements, their development seems founded more in brainstorming rather than a combination of brainstorming with a systematic SMS analysis. Also, the various interlinks between SMS activities are not yet explicitly addressed, and the several SMS parts are assessed individually.

This paper presents the AVAC-SMS metric designed for the self-assessment of the maturity of an aviation SMS and the extent to which the specific SMS (WaD) differs from the ideal SMS (WaI). The metric incorporates a tool that is customizable to the size and complexity of the organisation depending on the level of SMS assessment detail and results in quantified scores. Although the specific metric is not supposed to replace audit instruments currently used by the industry, its value compared to existing auditing tools lies on the fact that is based on a systematic analysis of SMS according to the Systems-Theoretic Process Analysis (STPA). The AVAC-SMS metric introduces the assessment of SMS process dependencies, which are not visible evaluated by current tools, and includes three distinct steps of SMS maturity assessment: institutionalisation, capability and effectiveness.

2   Methodology

2.1   Overall approach

The STPA technique (Leveson, 2011) comprised the basis for the tool development. The particular technique is originated in the Systems-Theoretic Accident Model and Processes (STAMP) that, unlike traditional safety models, embraces systems theory and thinking. Systems theory considers that degraded system performance might be a result of uncontrolled interactions among system components in addition to individual component failures and behaviours. STPA can be applied *a posteriori* to existing systems, or *a priori* to systems under development, and can lead to the identification of design errors, unsafe interactions between components, complexity factors in human decision-making as well as social, organisational and management factors contributing to poor performance.

STPA starts with the establishment of the system engineering foundation where the analyst defines the unwanted losses, system hazards and constraints and illustrates the basic hierarchical control structure of the system. The control structures show the responsible human or automated controllers, processes under control, actions, feedback mechanisms as well as information flow from other agents. Each controller manages the process based on a set point, which reflects the system objectives and constraints, a control algorithm, which denotes how the processes must be run, and through necessary actions. In parallel, the controller receives feedback from the process and information from other agents to update his/her/its process model, understand whether the process is run as expected and make any adjustments. This is how the gaps between WaI (setpoint and control algorithm) and WaD (system state and process performance) are minimised. It is clarified that in the glossary of STAMP, an accident is considered any loss that reflects the non-achievement of system goals (e.g., safety, security, efficiency, productivity). It is noted that this definition is broader than the ones used by the aviation industry (e.g., ICAO, 2001; EASA, 2010).

After the preparation phase explained above, STPA step 1 regards the examination of the cases that each control action could lead to a hazardous state if delivered when appropriate, if not delivered when necessary, if performed in wrong order or timeliness, or if applied too long/stopped too early. The list of Unsafe Control Actions (UCA) leads to the formulation of requirements for the controller as a means to maintain the system within the boundaries defined. In step 2, the causes that can lead to performing a UCA or rendering a 'safe' Control Action (CA) ineffective are examined. Leveson (2011) provides a list of keywords that assist the analyst in detecting

causal factors, and at this step, expert judgment is heavily required to conclude with a complete list of causes. STPA ends with the generation of causal scenarios that are the paths through which a system can fail. These scenarios are various combinations of causal factors that can lead to UCAs or ineffective CAs and are used to develop and carry out tests for evaluating system performance.

The researchers with the support of two undergraduate students (Abrini, 2016; Masoud, 2017) applied STPA on the aviation SMS as described by the Safety Management Manual of ICAO (2013). With the goal to provide a tool for SMS maturity assessment that combines compliance and performance without deviating from the vein of STPA and systems theory, we designed the AVAC-SMS tool based on the following reasoning:

- The requirements generated from UCAs reflect the institutionalisation of SMS: they include design and implementation points that also incorporate SMS process and time dependencies.
- The causal factors that can lead to UCAs can be used by the organisation to search for reasons that have led to the UCAs detected. This might lead to the identification of problems either internally to the organization or at higher levels (e.g., State or regional authorities) or even in the way the SMS is standardised internationally.
- The causes that might contribute to ineffective CAs reflect capability factors to run the SMS activities irrespectively of its institutionalisation level. On the one hand, an SMS can be correctly designed and implemented according to its design, but its deliverables can yet be unachievable. On the other hand, a low capability level can signal that an SMS might not be suitably tailored to the organisation.
- The quantity, quality and timeliness of SMS outputs can function as proxies for SMS effectiveness, the latter connected with the extent to which the outputs of SMS activities influence the execution of daily activities positively.

Furthermore, the researchers considered that a detailed SMS assessment might be feasible, even necessary, for large and complex companies but not for Small and Medium Enterprises (SME) that have limited resources and simpler SMS and organisational structures. Although all aviation companies are obliged or expected to implement an SMS, standards allow its scalability according to the size of the organisation and complexity of operations (e.g., FAA, 2015). Typically, SMEs follow the ICAO Safety Management Manual (ICAO, 2013) but properly tailored to their structure and activity volumes (e.g., parallel safety and other management tasks,

simplified risk management processes, short risk registries). Therefore, we aimed at maintaining the concepts of SMS institutionalisation, capability and effectiveness mentioned above, but allowing SMEs to assess these SMS dimensions through a less extensive instrument.

Moreover, the authors introduced quantified measurements and scoring of the various assessment areas to allow SMS maturity monitoring over time and internal or external benchmarking, hence contributing to a performance-based assessment through numerical figures. We used ratios for the degree to which each of the SMS requirements is met and for each of the capability and effectiveness parameters (i.e. values 0% to 100%). We calculate Euclidean distances across the whole SMS (i.e. all requirements/ factors as a single vector) or across SMS components and elements (i.e. the distances calculated within each component and element and then combined to a single SMS score) as a means to calculate the gap between WaI (SMS designed according to standards, fully implemented according to the SMS manual and with 100% effectiveness) and WaD (current SMS assessed).

2.2   Review and finalisation of the tool

All metrics developed in the frame of the research project were assessed against the set of the following criteria (Kaspers et al., 2016a) to ensure their accuracy, construct, content and face validity:

- reflective of the respective theoretical framework;
- encompassing systemic views, where applicable;
- valid (i.e. a meaningful representation of what is measured);
- fulfilment of laws, rules and other requirements, where applicable;
- measurable, so to permit statistical calculations;
- specific in what is measured;
- availability or easiness of obtaining hard or/and soft data required including the quantification of the latter;
- ability to set control limits for monitoring the calculated values;
- manageable – practical (i.e. comprehension of metrics by the ones who will use them);
- scalable/applicable to the context and area that the metric will be used (e.g., size of the company, type of activities such as air operations, maintenance, ground services, air traffic management);
- cost-effective, by considering the required resources;
- immune to manipulation;
- sensitive to changes in conditions.

To evaluate the fulfilment of the above criteria the researchers, after the draft design of the AVAC-SMS metric, subjected it to peer-reviews within the research team and with the engagement of knowledge experts (i.e. aviation authorities, universities, research institutions and consultants) and SME and large aviation companies (Table 1). Also, the underlying concept and the draft versions of the metrics were presented to four scientific and six industry conferences, where formative feedback was collected. All comments received by the reviewers and during the conferences spanned along various of the quality criteria mentioned above and led to the final design of the metric.

Table 1   Type and number of organisations participated in the reviewing process

Airlines	Air Navigation Service Providers	Ground Operations (Maintenance, Ground Handling, Airports)	Knowledge Experts
Round 1: April – June 2017			
6	1	1	3
Round 2: September – October 2017			
10	2	4	4

3   Results

In this section, we present the main results along with examples of the final SMS maturity assessment tool. It is noted that the full version of the tool and the control structures of the whole SMS and per component, which were based on their description by ICAO (2013), are available to the reader upon request to the corresponding author. For the scope of SMS assessment, we defined the following accidents, system hazards and system requirements according to the STAMP/STPA terminology.

System Accidents:

A-1. Level 1 SMS audit findings (i.e. the type of findings that indicate poor SMS design and implementation and can lead to suspension of the operating license).

A-2. Poor safety performance [i.e. under the premise that a well-operated SMS will increase safety performance (ICAO, 2013)].

Especially regarding system accident A-2, it is noted that, to date, indicators and thresholds of safety performance are defined differently by each company and agreed with the competent authority. Such indicators

typically include numbers and ratios of safety occurrences regarding operational and technical events (e.g., non-airworthy aircraft released for service). Also, SMS performance (e.g., timely response to an accident) is part of the safety assurance component of SMS and might be monitored by respective metrics, but it is not considered as part of safety performance.

System Hazards and Linkage with System Accidents:

    H-1. SMS is not designed according to standards (A-1, A-2)

    H-2. SMS is not implemented according to standards (A-1, A-2)

    H-3. SMS is not suitable for the organization (A-2)

    H-4. SMS is not effective (A-2)

Especially for the hazards H-1 and H-2, it is clarified that on the scope of this research we did not examine the completeness and quality of the SMS as described by ICAO (2013) and we assumed that the respective standards reflect the best SMS design (i.e. WaI). However, the causal factors linked to the occurrence of Unsafe Control Actions can reveal whether such standards lack feasibility, clarity, quality etc. and might indicate the need for corresponding changes.

System-level Requirements:

    SR-1: SMS shall be designed according to standards

    SR-2: SMS shall be implemented according to standards

    SR-3: SMS shall be suitable for the organization

    SR-4: SMS shall be effective

3.1 SMS institutionalisation

In overall, the assessment of the extent to which an SMS has been institutionalised is performed by the safety department according to the options described below in sections 3.1.1 and 3.1.2. The assessment can be based preferably on available records as well as information from internal and external audits, studies etc.

3.1.1 Full-scale SMS institutionalisation assessment

The application of STPA in conjunction with the analysis iterations following the remarks of the reviewers led to 149 UCAs and corresponding requirements for the design and implementation of an SMS. Table 2 presents the distribution of these requirements across the SMS components and elements.

Table 2 Distribution of SMS requirements

SMS Component	SMS Element	Number of SMS Requirements			Total	
		Design	Implementation	Dependencies	Element	Component
Safety Policy & Objectives	Management Commitment and Responsibility	13	8	9	30	73
	Accountabilities and Responsibilities	2	2	4	8	
	Assignment of Resources & Appointment of Key Personnel	5	5	1	11	
	Coordination of Emergency Response Planning	6	3	4	13	
	SMS Documentation	4	4	3	11	
Safety Risk Management	Hazard Identification	4	4	3	11	23
	Risk Assessment and Mitigation	2	5	5	12	
Safety Assurance	Safety Performance Monitoring & Measurement	7	6	7	20	36
	The Management of Change	2	3	5	10	
	Continuous Improvement of SMS	1	2	3	6	
Safety Promotion	Training & Education	2	3	3	8	17
	Safety Communication	3	5	1	9	
Total		51	50	48	149	149

Table 3 illustrates an extract of the set of requirements for the SMS element Management Commitment and Responsibility coded as MCR. For illustration purposes, the requirements in Table 3 are differently coloured per type (design: blue, implementation: orange, dependency: green). Each checking point can get scores between 0% - 100% with increments of 20%. Where applicable, the design requirements can only get values of 0% or 100% due to their binary nature. The specific score method was preferred to ease the user and avoid mistyping. The total SMS institutionalisation score is calculated by measuring the differences between the ideal SMS (i.e. all requirements fully met) and its current state (i.e. actual scores).

Table 3   Example of SMS Institutionalisation Requirements (blue: design/ documentation, orange: implementation, green: dependencies).

Code	Check if:	Extent of Realising the Requirement
MCR1	There is a safety policy	0%
MCR2	The overall organisational policy views safety as one of the core business functions	0%
MCR3	Safety staff and officers participate in all planning and review management meetings (across all organizational levels and sections, as applicable)	0%
MCR4	Safety is a parameter in decision-making during all planning and review management meetings (across all organizational levels and sections, as applicable)	0%
MCR5	The possible need to change the safety policy has been always discussed by management during significant changes within the organization	0%
MCR6	Current safety policy is included in all safety education/training programs	0%

3.1.2   Middle-scale and Short-scale SMS institutionalisation assessment

To accommodate the need for SMS assessment in SMEs, the metric offers the option to evaluate the SMS as a whole or at the levels of elements and components by using 4 question types for each case: design according to standards, implementation according to standards, accomplishment of activities within defined timelines, and exploitation of inputs from other SMS and organizational activities. In this case, instead of 149 checking points (see section 3.1.1), the companies can opt in assessing their SMS with the use of either 4 questions (whole SMS) or 16 questions (SMS components) or 48 questions (SMS elements). However, it is noted that the higher the assessment detail, meaning the lower the resolution, the more the insights

to be gained by the company about its SMS and the higher the validity and reliability of the results. The scoring concept, in this case, is the same one applied to the detailed SMS institutionalisation assessment. For example, the questions asked for each of the SMS elements are:

- To what degree are the activities included in this element designed/documented according to standards?
- To what degree are the activities of this element implemented as described in standards?
- To what degree are the activities of this element accomplished within the defined timelines / when needed?
- To what degree are the activities of this element performed by using necessary information from other SMS and organisational activities?

3.2   SMS capability

The assessment of SMS capability is achieved through surveys to all managers and safety personnel who are responsible for operationalising the SMS. On this scope, six types of questions were introduced according to the keywords used by STPA in the analysis of causes leading to ineffective control actions: capability of the controller (i.e. the person responsible to implement the SMS tasks), adequacy of means provided, degree of conflict with other controllers, adequacy of inputs from other organizational/SMS activities, timely delivery of such inputs, and degree of influence of external disturbances. It is noticed that the "adequacy" term is used to reflect both quantity and quality of the means provided or inputs used. The researchers did not decompose the specific assessment topic as a means to avoid overwhelming the participants with many questions.

Since answering to six questions for each of the 149 SMS checking points (section 3.1.1 above) would be impractical, the questions can be posed for each of the SMS elements or components or the whole SMS, depending on the resolution level preferred. For instance, the following questions are asked for each SMS element:

- How capable do you feel of executing your tasks related to this element?
- How adequate are the means available to you to execute the tasks related to this element?
- To what degree do you conflict with other persons that work on the same tasks of the SMS element?
- How adequate is the information from other organisational and SMS activities you need to execute the tasks of this element?
- How timely do you receive necessary information from other organisational and SMS activities to execute your tasks of this element?

- To what degree do external factors disturb you in the execution of your tasks of this element?

Each of the survey questions is answered based on a scale between 0% and 100% with increments of 20%. The overall capability per point assessed is the average of the scores of the six questions across the whole sample. The final SMS capability score is calculated from the difference between the ratings and the ideal SMS, where the ideal SMS is the one where the overall capability is 1.0.

3.3   SMS effectiveness

The SMS effectiveness is evaluated through three questions that can be posed for the whole SMS or per component and element. These questions will be addressed to all employees who receive the final deliverables of the various SMS activities as a means to perform their tasks safely while achieving the rest of their objectives (e.g., quality and timely delivery). The three questions target at evaluating the quantity, quality and timeliness of SMS activities and a scale between 0% and 100% with increments of 20% is used. The overall SMS effectiveness score is calculated as with the case of SMS capability assessment. For instance, the questions asked for the effectiveness of SMS elements are the following:

- To what degree is the amount of activities related to this element adequate to support your daily tasks?
- To what degree is the quality of the activities related to this element sufficient to support your daily tasks?
- How timely are the activities related to this element executed to support your daily tasks?

3.4   Causal factors for low SMS institutionalisation level

In case that the score of institutionalisation for the whole SMS or for particular elements, components or checking points is lower than the organisation expects or tolerates, an additional survey can be administered to managers and/or safety staff involved in the SMS design and implementation. The research team generated a list of causal factors based on the instructions of the STPA technique, but it was outside of the scope of this study to suggest a relevant survey instrument. This extra assessment step regards factors that are not viewed as part of the SMS institutionalisation but can explain SMS deficiencies and allow organisations to devise remedies or suggest improvements at higher hierarchical levels (e.g., standardisation

agencies, State and regional authorities). For example, the following causal factors apply to the SMS element assessment level:

- The SMS element is documented inadequately/poorly (e.g., clarity, accuracy).
- The SMS element activities are not sufficiently known.
- Inadequate information and feedback required to perform the activities of the SMS element.
- Information and feedback required to perform the activities of the SMS element are provided with a delay.
- Information and feedback needed to perform the activities of the SMS element are corrupted or of poor quality.

3.5   Total SMS maturity level

The results from the assessment of SMS institutionalisation, capability and effectiveness can be combined to provide the overall SMS maturity level of the organisation. To align with the industrial practice that views SMS maturity as a ladder (e.g., SMICG, 2012; Eurocontrol, 2012), the researchers suggest the addition of the results of each of the SMS assessment areas. Since each of these areas can get a maximum score of 1.0, the maximum SMS maturity level can be 3.0. It is noted that apart from the overall SMS score, the metric can provide maturity scores per SMS element and component, thus offering companies the ability to prioritise their SMS improvement initiatives. The definition of the thresholds of SMS maturity scores that can signal poor, adequate or excellent SMS performance was out of the scope of this research, but they can be defined initially by each company and later standardised across the aviation industry.

4   Discussion

The SMS assessment metric/tool presented in this paper was developed based on the Safety Management Manual of ICAO (2013) and the System Theoretic Process Analysis (STPA) technique (Leveson, 2011). The specific metric incorporates the view of SMS as a system by addressing its institutionalisation (i.e. design and implementation along with time and internal/external process dependencies), capability and effectiveness. Each of these assessment areas leads to individual scores, which in combination they provide the total SMS maturity score. Particularly and in connection with the quality criteria used in our study, the Appendix shows the remarks of the researchers that address the various comments and concerns stated by the tool reviewers during its development.

It is clarified that an SMS assessment with the use of the suggested metric can be viewed as a starting point. Depending on the results of SMS self-assessments, organisations can proceed to a collection of qualitative data with a focus on the weakest areas revealed by the initial assessment. The indicative list of causal factors provided in section 3.4 above can function as guidance to design specific interview or questionnaire items. Moreover, the scores of each SMS dimension and per SMS component and element can be examined further to detect differences among organizational levels and functions and indicate areas where the gaps between WaI and WaD are higher and necessitate interventions with higher priority.

Regarding the differences between the proposed metric and existing instruments, such as the ones developed by Eurocontrol (2012), SMICG (2012) and EASA (2017), the AVAC-SMS tool was based on STPA that provides a consistent and systematic manner for assessing a system without excluding the value of expert judgment and staff perceptions. The AVAC-SMS metric (1) includes dependencies, which are not explicitly addressed in current tools, (2) assesses the SMS capability as proxy for the SMS suitability, which cannot be evaluated through existing tools due to the lack of respective instructions, and (3) employs a specific set of questions as proxies for the SMS effectiveness, whereas current tools attempt to evaluate the latter through questions formulated based mostly on experience.

The detail of assessment concerned, the metric offers different options depending on the resources each organisation plans to invest in SMS assessment (Table 4). However, whereas the long-scale SMS assessment is viewed as sufficiently valid and reliable, these characteristics for the short and medium scale assessments will be tested in the next research phase through the application of the metric to companies.

Table 4   Possible SMS Assessment Resolutions

Survey Questions to Staff: 1. SMS Capability (Managers/Safety Personnel) 2. SMS Effectiveness (Employees)	Number of Assessment Points (Deskwork): SMS Institutionalisation		
	149 (SMS Processes)	48 (SMS Elements)	16 (SMS Components)
Questions at the level of SMS elements 1. *72 Capability questions* 2. *36 Effectiveness questions*	Option SE	Option EE	Option CE
Questions at the level of SMS components 1. *24 Capability questions* 2. *12 Effectiveness questions*	Option SC	Option EC	Option CC
Questions about the whole SMS 1. *6 Capability questions* 2. *3 Effectiveness questions*	Option SS	Option ES	Option CS

5 Conclusions

The metric designed for the self-assessment of SMS maturity fills the gaps of existing tools but is not meant to replace formal audits. It is supposed to complement current SMS assessment tools used in audits and enable organisations to perform a systematic evaluation of their SMS to the extent desired and detect strong and weak areas. Following two review rounds, the researchers developed the metric by ensuring that it meets the quality criteria mentioned in literature to the maximum extent possible. Also, the particular metric meets sufficiently the need of SMEs for safety metrics that do not obtain vast amounts of operational data.

It is envisaged that the metric satisfies the requirements for a performance-based assessment and it is uniform in the sense that it can be used by any aviation organization/service provider with an established ICAO-based SMS. Also, the AVAC-SMS metric can be appropriately adapted by other industry sectors to accommodate any other type of safety management system. Overall, the SMS assessment tool suggested in this paper aims at assisting organisations with the assessment of their SMS and the scores generated can enable the qualitative and numerical monitoring of SMS maturity and the quantification of the effects of planned or implemented SMS changes. Apart from improving SMS, the tool can be used when SMS and safety performance targets are not met or are not within defined limits/thresholds.

The AVAC-SMS tool adequately meets the accuracy, construct, content and face validity types, thus suggesting its potential to be immediately used by companies regardless of the results of criterion, predictive, and statistical conclusion validity at the next research phase. The next step will be the application of the AVAC-SMS metric to different companies, collection and analysis of data and examination of the associations of SMS institutionalisation, capability, effectiveness and overall maturity with safety/system outcomes.

Acknowledgments

Hol (Transavia, NL), Dimitrios Karampelias (Athens International Airport, GR), Keith Cartmale Keith (MUAC, NL), Charlotte Mills (New Zealand CAA, NZ), Conor Nolan (Aer Lingus, IE), John John (Av8jet, UK), Mark Overijnder (Jetsupport, NL), Ilias Panagopoulos (NATO Support and Procurement Agency, HU), Anastasios Plioutsias (National Technical University of Athens, GR), Gert Jan van Hilten (KLM Ground Services, NL), Jaco vd Westhuizen (ATNS, ZA).

Appendix

Quality Criteria (Kaspers et al., 2016a)	Remarks
Reflective of the respective theoretical framework	The metric is used to reveal the distances between WaI and WaD through the application of STPA, which is a published and broadly applied analysis technique. The SMS assessment allows the use of hard SMS data (e.g., reports and records) for the SMS institutionalisation dimension and perceptions of personnel from all organisational levels in the assessment of SMS capability (i.e. managers/safety staff) and effectiveness (i.e. all employees). Under this approach, the SMS assessment is achieved by including all persons involved in the design and operation of SMS along with the recipients of its deliverables.
Encompassing systemic views	The metric addresses all sides of SMS maturity under a systems approach: design, implementation, dependencies, capability and effectiveness of outputs.
Valid (i.e. meaningful representation of what is measured)	The validity of content and analysis was ensured by adhering to the ICAO SMS standard and by ensuring the consistent application of STPA. The validity of the metric will be examined further regarding the (1) relationships between the various SMS scores with safety performance, and (2) reliability of different levels of SMS assessment (i.e. all checkpoints, elements, components or the whole SMS).
Fulfilment of laws, rules and other requirements, where applicable	The metric includes the principal points of the ICAO Safety Management Manual properly enriched according to the comments of the reviewers. The metric can be modified to include more detailed SMS tasks or new/amended ones in case the standards change.
Measurable, so to permit statistical calculations	The metric concludes with quantified results.
Specific in what is measured	Each assessment step explicitly measures different dimensions of SMS
Availability or easiness of obtaining hard or/and soft data required including the quantification of the latter	The metric does not mandate the use of hard data for the SMS institutionalisation assessment, but this is preferred. The choice will depend on the organisation. The SMS capability and effectiveness are assessed through surveys to employees to whom a time slot can be given in their schedule to participate.

Quality Criteria (Kaspers et al., 2016a)	Remarks
	The metric will be converted to an online platform to support the easier collection of data.
Ability to set control limits for monitoring the calculated values	Each organisation can set limits for the SMS institutionalisation, capability and effectiveness in overall or per component, element and activity.
Manageable – practical (i.e. comprehension of metrics by the ones who will use them)	The SMS institutionalisation assessment requires a good knowledge of SMS by the safety staff who will use it.
	The SMS capability and effectiveness dimensions require a basic knowledge of the activities included in each SMS component and element. The metric provides a brief description of each component and element as a means to help employees to recall the corresponding SMS activities. Nonetheless, basic knowledge about SMS is provided to staff through mandatory training and education.
Scalable/applicable to the context and area that the metric will be used (e.g., size of the company, type of activities such as air operations, maintenance, ground services, air traffic management)	The metric can be used by any type of organization because it does not refer to specific operational activities.
	The metric provides different resolutions for the SMS assessment to accommodate companies of different sizes and complexity. The reliability of the middle and short scale assessments will be examined in the next research phase.
Cost-effective, by considering the required resources	The four different assessment levels reflect different resource demands. It is envisaged that large and complex organisations will use the detailed assessment steps and SMEs will use the middle and short scale assessments, pending the examination of the reliability of the latter.
	The SMS self-assessment tool, especially in its detailed form, is supposed to be used with a long periodicity (e.g., once a year) or whenever organisations are concerned about their safety performance or receive significant audit findings.
	An adequate and representative sample will suffice to reach to valid results. However, the larger the sample and the investment of resources, the higher the validity.
Immune to manipulation	The reliability of SMS institutionalisation assessment will depend on the choice of the organisation to use logs and documented data or trust the memory of safety staff.
	The effects of socially desirable answers can be avoided by a clear communication about the goal of the assessment and the anonymity of the survey participants.
	The metric is supposed to be used for self-assessment and does not comprise an audit tool. Therefore, honest answers will be expected.
Sensitive to changes in conditions	The periodical application of the metric can capture differences due to SMS and organizational changes, changes in workforce diversity etc. The periodicity depends on the company.
	The use of distance calculations and not merely differences/ratios increases the sensitivity of SMS scores across all assessment areas.

References

Abrini, M. (2016). *STPA-based SMS Assessment Tool* (Unpublished Bachelor Thesis). The Netherlands: Aviation Academy, Faculty of Technology Amsterdam University of Applied Sciences.

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

Dekker, S. (2011). *Drift into Failure: From Hunting Broken Components to Understanding Complex Systems*. Farnham, UK: Ashgate Publishing.

EASA (2010). *Regulation (EU) No 996/2010 of the European Parliament and of the Council of 20 October 2010 on the investigation and prevention of accidents and incidents in civil aviation and repealing Directive 94/56/EC*. Cologne: European Aviation Safety Agency.

EASA (2014). *A Harmonised European Approach to a Performance-Based Environment*. Cologne: European Aviation Safety Agency.

EASA (2017). *Management System Assessment Tool*. Cologne: European Aviation Safety Agency.

Eurocontrol (2012). *Effectiveness of Safety Management*. Brussels: Eurocontrol.

FAA (2006). *Introduction to Safety Management Systems for Air Operators. Advisory Circular 120-92*. USA: Federal Aviation Administration.

FAA (2015). *Safety Management Systems for Domestic, Flag, and Supplemental Operations*. USA: Federal Aviation Administration.

ICAO (2001). *Aircraft Accident and Incident Investigation, Annex 13 to the Convention on International Civil Aviation*. (9th ed.). Montréal, Canada: International Civil Aviation Organization.

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

ICAO (2016). *Safety Management, Annex 19 to the Convention on International Civil Aviation* (2nd ed.). Montréal, Canada: International Civil Aviation Organization.

Karanikas, N. (2016). Critical Review of Safety Performance Metrics, *International Journal of Business Performance Management*, *17*(3), 266-285.

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

Kaspers, S., Karanikas, N., Roelen, A.L.C., Piric, S., van Aalst, R., de Boer, R. J. (2016b). *Results from Surveys about Existing Aviation Safety Metrics, RAAK PRO Project: Measuring Safety in Aviation, Project S10931*. The Netherlands: Aviation Academy, Amsterdam University of Applied Sciences.

Kaspers, S., Karanikas, N., Roelen, A., Piric, S., van Aalst, R., de Boer, R. J. (2016c). Exploring the Diversity in Safety Measurement Practices: Empirical Results from Aviation. Proceedings of the 1st International Cross-industry Safety Conference, Amsterdam, 3-4 November 2016, *Journal of Safety Studies*, *2*(2), 18-29.

Leveson, N. (2011). *Engineering a safer world: Systems thinking applied to safety*. Boston, Mass: MIT Press.

Leveson, N. (2015). A systems approach to risk management through leading safety indicators. *Reliab Eng Syst Saf*, *136*, 17–34.

Masoud, S. (2017). *SMS Tool: Quantitative Evaluation of Safety Management Systems* (Unpublished Bachelor Thesis). The Netherlands: Aviation Academy, Faculty of Technology Amsterdam University of Applied Sciences.

SMICG (2012). *Safety Management System Evaluation Tool*. Safety Management International Collaboration Group.