RAAK PRO Project: Measuring Safety in Aviation

Deliverable: Finalisation and Application of New Safety Management Metrics

December 2018

Nektarios Karanikas, Alfred Roelen, Alistair Vardy and Steffen Kaspers

Project number: S10931

RAAK PRO Project: Measuring Safety in Aviation

Finalisation and Application of New Metrics

Nektarios Karanikas[1], Alfred Roelen[1,2], Alistair Vardy[1], Steffen Kaspers[1]

[1]Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands
[2]NLR, Amsterdam, the Netherlands

Contents

Executive Summary

Following the completion of the 2nd research phase regarding the design of new safety metrics that could be used in Safety Management Systems (SMS), Section 2 of this report explains the methodology of designing the five new metrics: the AVAC-SMS for the self-assessment of Safety Management Systems, the AVAC-SCP for the assessment of Safety Culture Prerequisites (SCP) that companies could plan and implement to foster a positive safety culture, three indicators for assessing the effectiveness of risk controls, five indicators reflecting the utilization of organisational resources, and a metric for the complexity of socio-technical systems. Section 3 presents briefly the particular metrics which have been published as part of the proceedings of the 2nd International Cross-industry Safety Conference (Amsterdam, 1-3 November 2017). Section 4 of the report discusses the application of two of the metrics by companies (i.e. AVAC-SMS and AVAC-SCP), and section 5 presents the respective results. The report concludes with a discussion of the results and suggestions for the next project steps.

Overall, the application of the metrics showed that they have adequate sensitivity to capture any gaps between Work-as-Imagined and Work-as-Done amongst different organizational levels and across organizations. Also, the results revealed interesting differences between the various areas measured with each metric: Institutionalization, Capability and Effectiveness for the AVAC-SMS, and Planning, Implementation and Perceptions for the AVAC-SCP. However, the relatively small sample of companies and restricted number of managers and employees participating in each company render the findings only indicative and not conclusive. Also, this limitation did not allow to perform comparisons between large companies and SMEs as well as amongst companies with different operational activities (i.e. airlines, air navigation service providers, airports and ground services).

At this stage, due to the limited size and composition of the sample and the few safety/activity data provided my companies we could not determine whether the metrics have any predictive validity. The researchers plan to run a second round of surveys to apply the metrics and collect safety/activity data from more organizations, hence we anticipate that we will be able to test the metrics against safety performance and activity figures. Nonetheless, irrespective of the possible associations of the metrics with safety outcomes, their application and findings communicated in this report are supportive of their usefulness, practicality and potential value for the companies that are interested in assessing their SMS and SCP, reveal gaps amongst the specific assessment areas per metric and get insights into their strong and weak points to improve further the way they manage safety.

1. Introduction

In September 2015, the Aviation Academy of the Amsterdam University of Applied Sciences initiated the research project entitled “Measuring Safety in Aviation – Developing Metrics for Safety Management Systems” which is co-funded by the Regieorgaan Praktijkgericht Onderzoek SIA[1]. The project responds to the specific needs of the aviation industry: Small and Medium Enterprises (SME) lack large amounts of safety-related data to measure and demonstrate their safety performance proactively; large companies might obtain abundant data, but they need safety metrics which are more leading than the current ones and of better quality; the transition from compliance-based to performance-based evaluations of safety is not yet backed with specific tools and techniques. Therefore, the research aimed to identify ways to measure safety proactively in scientifically rigorous, meaningful and practical ways without the benefit of large amounts of data and with an emphasis on performance rather than mere compliance (Aviation Academy, 2014). During the first phase of the project, the research concluded to the findings and design concepts briefly described in the following paragraphs.

State-of-art academic literature, (aviation) industry practice, and documentation published by regulatory and international aviation bodies jointly suggest that (a) safety is widely seen as avoidance of failures and is managed through the typical risk management cycle, (b) safety metrics can be, conventionally, split in two groups: safety process metrics and outcome metrics, (c) the thresholds between the different severity classes of safety occurrences are ambiguous, especially between incidents and serious incidents, (d) there is a lack of standardization across the aviation industry regarding the development of safety metrics and the use of specific quality criteria for their design, (e) safety culture is seen as either a result of safety management or a reflection and indication of safety management performance), and (f) there is limited empirical evidence about

[1] http://www.regieorgaan-sia.nl/

the relationship between Safety Management System(SMS)/safety process and outcome metrics, and the link between those often relies on credible reasoning (Karanikas et al., 2016b; Kaspers et al., in press).

Initial results from surveys conducted to 13 aviation companies (i.e. 7 airlines, 2 air navigation service providers and 4 maintenance/ground service organizations) showed that (a) current safety metrics are not grounded on sound theoretical frameworks and, in general, do not fulfil the quality criteria proposed in literature, (b) safety culture is not a consistent part of safety metrics and, therefore, not assessed, (c) companies collect data related to their SMS processes, but such data are not associated with SMS metrics, (d) the safety management-related data in use differ across companies depending on own perceptions, safety models adopted implicitly or explicitly, and available resources, (e) SMS assessment is yet based on a compliance-based approach, (f) a few, diverse and occasionally contradictory monotonic relationships exist between SMS process and outcome metrics. The latter finding was attributed to a combination of factors, which are linked to the limitations of a linear approach and the different ways SMS processes are implemented, and safety outcomes are classified (Karanikas et al., 2016a; Kaspers et al., 2016, 2017).

Taking into account the current situation and after reviewing relevant literature (Karanikas et al., 2017a), the research team contemplated that the gaps between work as prescribed in rules and procedures (a.k.a Work as Imagined – WaI) and work as actually performed (a.k.a. Work as Done - WaD had not been sufficiently and evidently illustrated through relevant metrics. Thus, the primary focus of the researchers was the distance between WaI and WaD, under the suggestion that if those get close, the changes can be induced to both or either of them. Only the gaps were of interest, and the authors did not suggest either WaI or WaD as more or less appropriate for achieving the system objectives, because this requires deep knowledge of each context, which was out of the scope of the particular research. To develop new safety metrics, the researchers initially reviewed relevant literature to identify how the WaI-WaD gaps could be depicted and quantified. The safety metrics that were perceived as suitable to be operationalized through respective metrics were (1) SMS self-assessment based on the System-Theoretic Process Analysis, (2) Safety Culture Prerequisites assessment that complements Safety Culture assessments, (3) effectiveness of risk controls, (4) the distance between WaI and WaD at the operational level, (5) complexity measurement of a socio-technical system, and (6) utilisation of resources (Karanikas et al., 2017b). It is noticed that the metric regarding the effects of the WaI-WaD gaps on safety performance is part of PhD research at the Delft University of Technology which is conducted by a research team member. The particular research is expected to conclude by the end of this project and retrofit the overall results. Therefore, the rest of this document regards the other five metrics.

2. Methodology

The criteria against which accuracy, construct, content and face validity of the different versions of the metrics were assessed are the following [adapted from Karanikas et al. (2017) and Kaspers et al. (in press) and addressing the limitations of current metrics presented in section 1 above]:

- reflective of the respective theoretical framework;
- encompassing systemic views, where applicable;
- valid (i.e. 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 metrics, subjected those 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's and large aviation companies (Table 1). The distribution of the organizations that reviewed the metrics in each round was decided

by considering the maturity level and length of each of the metrics and the availability of the reviewers. Also, the underlying concepts and the draft 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 metrics.

Review rounds and metrics	Airlines	Air Navigation Service Providers	Ground Operations (maintenance, ground handling, airports)	Knowledge Experts
Round 1: April – June 2017				
SMS assessment tool	6	1	1	3
SCP tool	2	1	1	4
Complexity/coupling	2	2	-	-
Risk control effectiveness	2	-	-	1
Resource gaps	3	-	-	2
Round 2: September – October 2017				
SMS assessment tool	10	2	4	4
SCP tool	10	2	4	2
Complexity/coupling	9	1	-	2
Risk control effectiveness	10	-	5	2
Resource gaps	10	-	5	2

Table 1: Reviews of metrics (numbers of participating organisations/companies)

The internal and external reviews of the metrics resulted in their finalisation. The concept, objective and design of each metric were presented at the 2nd International Cross-industry Safety Conference and published in the conference proceedings. In the following section, we describe the metrics briefly along with the corresponding references for the convenience of the reader.

3. Brief Description of Metrics

3.1 SMS assessment (Karanikas et al., 2018)

The Aviation Academy SMS assessment metric/tool (named as AVAC-SMS) 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 the areas of institutionalisation (i.e. design and implementation along with time and internal/external process dependencies), capability (i.e. to what extent managers have the capability to implement the SMS) and effectiveness (i.e. to what extent the SMS deliverables add value to the daily tasks of employees). The assessment of each of these assessment areas leads to individual scores which can illustrate the gaps between them.

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. Moreover, the scores of each SMS area and per SMS component and element can be examined further to detect differences amongst 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 based on the three principal traits of process deliverables (i.e.

quantity, quality and timeliness), 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. The list mentioned below is in descending order of detail:

- SMS institutionalisation (Safety Department). SMS tasks/processes level: 149 questions; SMS elements level: 48 questions; SMS components level: 16 questions.
- SMS capability (Managers). SMS elements level: 72 questions; SMS components level: 24 questions, Overall SMS level: 6 questions
- SMS effectiveness (Frontline Employees). SMS elements level: 36 questions; SMS components level: 12 questions, Overall SMS level: 3 questions.

However, whereas the longer SMS assessment can be expected as sufficiently valid and reliable (i.e. SMS institutionalisation at the task level and SMS capability and effectiveness at the element level), these characteristics for the short and medium scale assessments were tested through the application of the metric to companies, as explained in the respective section below.

The metric designed for the self-assessment of SMS 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. 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.

3.2 Safety Culture Prerequisites metric (Piric et al., 2018)

The researchers developed the Aviation Academy Safety Culture Prerequisites tool (named as AVAC-SCP) which was based on a previously published framework (Karanikas et al., 2016c) and combined 37 prerequisites to foster a positive safety culture. The prerequisites are clustered in six categories following Reason's (1998) typology of safety culture (i.e. just, flexible, reporting, informative and learning sub-cultures) and one additional category named general organisational prerequisites. The original objective of the tool was to gain insights into what prerequisites an organisation has included in their safety plans and to what degree the organisation safety culture plans are operationalised. Each of the prerequisites was transformed into questions to be answered by (1) safety managers who must check the organisational documentation to detect whether each prerequisite is present, and (2) safety and line managers regarding the implementation of the corresponding prerequisite.

However, the added value of the perception of safety culture aspects by the workforce could not be neglected; regardless of the efforts of a company to foster a positive safety culture, the perception of the workforce might differ from the intended outcomes of implemented plans. Therefore, in its final version, the AVAC-SCP was complemented with ten questions used to capture the perception of the employees and based on a condensed version of an existing safety culture assessment tool (NLR, 2016). The selection of only ten perception questions followed the advice given during the peer-review of the specific metric to decrease the number of questions addressed to frontline staff as a means to minimise the time needed to fill in the questionnaire and avoid boredom, tiredness or socially desirable answers when responding. Figure 1 shows a visual representation of the three elements in the tool.

Each assessment area results to an overall score which is used to evaluate the gaps between planning, implementation and perception, which, in turn, reflect the gaps between Work-as-Done and Work-as-Imagined at two different levels (i.e. safety department – managers, and managers-employees).

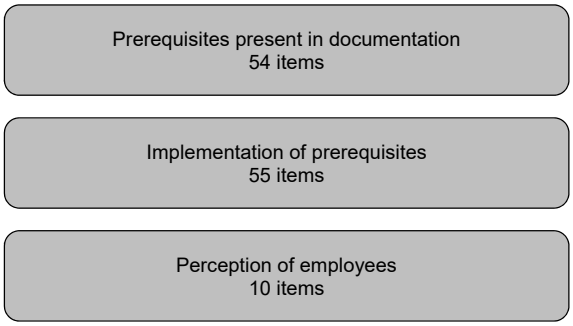

Figure 1: The structure of the AVAC-SCP tool

3.3 Effectiveness of risk controls (Roelen et al., 2018a)

The definition of effectiveness is "the degree to which something is successful in producing the desired outcome" (OED, 2017). In other words, the effectiveness of a risk control provides information on how many times the risk control is addressed in tackling a particular hazard or risk and how many of these times the risk control performs according to the desired outcome of the specific risk control. A generic indicator is developed based on this definition of effectiveness (Muns, 2017):

The ratio between the number of times a risk control is challenged and the amount of times the risk control achieves a successful[2] outcome.

Based on the definition above, the effectiveness of a risk control provides information on how many times the risk control is addressed in tackling a particular hazard or risk and in how many of these cases the risk control performs successfully. The following metrics were developed to determine the performance of risk controls:

$$1 - \frac{number\ of\ failures\ of\ the\ control\ when\ challenged}{number\ of\ occasions\ the\ control\ was\ challenged} \quad (1)$$

$$1 - \frac{number\ of\ failures\ of\ the\ control\ when\ tested}{number\ of\ occasions\ the\ control\ was\ tested} \quad (2)$$

$$1 - \frac{number\ of\ unwanted\ events\ after\ control\ was\ implemented\ per\ unit\ of\ time}{number\ of\ unwanted\ events\ before\ control\ was\ implemented\ per\ unit\ of\ time} \quad (3)$$

These metrics are listed in preferential order with the most preferred on top. A failure of risk control is defined as a failure to result in the specific desired outcome of the specific risk control. Because for some risk controls it may not be possible to observe if it is challenged, equations 2 and 3 are provided. Equation 2 relates to dedicated tests of the risk control (e.g. testing of the fire alarm during a fire drill), while equation 3 compares situations before and after implementation of risk control. For all three metrics, it is necessary to have an unambiguous description of the risk control as well as a description of the hazards(s) that the risk control must mitigate. It is also necessary to define what constitutes a failure of the risk control. The steps suggested to implement the metrics are: Describe the risk control; Determine how to identify a failure of the risk control; Determine whether it is possible to identify a challenge to the risk control (i.e. when the control was required to operate in real cases); Determine whether it is possible to test the risk control; Select a suitable time period; Collect data; Calculate risk control effectiveness.

3.4 Complexity of socio-technical system (Van Aalst et al., 2018)

The complexity metric was based on a review of the corresponding literature (see the full paper) which concluded to two complexity dimensions: the system complexity and perceived complexity. The former refers to the design and dynamics of system elements and interactions, and the latter is connected with the characteristics of human performance. This distinction was necessary since identical systems can be perceived more or less complex by various users. The parameters used for the formula of overall complexity (see below) for a given system are the number of system elements (NE), the number of elements interacting

[2] Successful is according to the specific desired outcome of the specific risk control.

with element *i* (NI*i*), the rate of distance change between elements *i* and *j* (d*ij*) that reflects the time window to react, the system slacks (SL) referring to the availability of Human Resources, Technical Resources and Communication/Coordination to control the system, and the four control modes as defined by Hollnagel (2017), being 4=scrambled or random control mode, where the operator has no idea what to do and acts impulsively, 3=opportunistic mode, 2=tactical mode and 1=strategic control mode (i.e. HP in the formula). The particular modes are seen as the result of parameters such as task difficulty, task load and workload.

$$SC_{perceived} = SC * HP = \left[\sum_{i=1}^{NE}\sum_{j=1}^{NI_i}\left(\frac{-\dot{d}_{ij}}{d_{ij}}\right)\right] * \frac{1}{SL} * HP$$

3.5 Utilisation of resources (Roelen et al., 2018b)

The specific metric considered four types of resources: Time, People, Money and Equipment. The defined indicators of resources utilization are:

- Available runtime / required runtime

Runtime is the turnaround time of a task and is a measure of task duration. Available run-time is the runtime that is scheduled for a specific task or group of tasks. Required runtime is the time that was actually needed to perform the task. Generally speaking, if the required runtime is longer than the available, there will be some sort of delay. If the required runtime is shorter than the available runtime, there is some sort of slack. Available short-term runtime is usually determined during activity planning and can be found in planning documentation. Actual run-time can be found in operational service records.

- Available person hours / required person hours

Person hours are a measure of total task effort. Available person-hours are the number of person-hours that are scheduled for a specific task or group of tasks. Required person-hours are the number of person-hours that were actually needed to perform the task. Generally speaking, if the required person-hours are more than the available person-hours, there will be a delay unless additional staff are made available. Available person-hours are usually determined during activity planning and can be found in planning documentation. Actual person-hours can be found in operational service records.

- Voluntary staff turnover

Voluntary staff turnover is defined as the percentage of employees in a workforce that voluntarily leave the organisation during a certain period of time (e.g., one year). Voluntary staff turnover data are usually recorded by the Human Resources department.

- Budget invested / budget spent

Budget invested/spent should be calculated for a specific activity or group of activities in a certain time period. Information on invested and spent budget can usually be found in the finance department.

- Number of equipment available / number of equipment required

Equipment available refers to the number of equipment that is actually available to perform the task under consideration. To be available, the equipment must be in working condition. Equipment required refers to the number of equipment that is actually required to perform a task. If the equipment required is less than the equipment available there is a shortage of equipment.

It is noted that available/required runtime and person-hours are to be calculated for a task or a combination of tasks depending on the system under study. Also, in the indicators above that are calculated through ratios, the nominator corresponds to the WaD and the denominator to the WaI. They can be calculated over different time periods (daily, weekly, monthly etc.) depending on the resources and focus of the organisation. Also, the metrics collectively provide broad coverage while being individually rather specific. This is considered a desirable attribute of performance indicators (Fitzgerald et al. 2011). The metrics described here are considered to be pro-active indicators, using the definition of Rasmussen et al. (2000) who defined pro-active indicators as indicators before an accident. Moreover, they are considered predictive in the sense

that they are predictive of the likelihood of occurrence of unsafe events, as opposed to monitoring indicators that use actual events as a measure for the likelihood of unsafe events (Körvers, 2004). It has been noted that there may be interactions among these indicators, for example, that voluntary turnover could affect financial performance (Shaw et al., 2005) and hence that there is a relation between the 'people' and 'money' indicators'.

4. Application of new safety metrics

4.1 Exclusion, inclusion and conversion of safety metrics

Following the finalisation of the metrics presented above and considering further feedback received from the project partners and during the interim project review (September 2017), the research team contemplated which metrics could serve better the objectives of the research and could be applied to aviation companies to collect and analyse data. The peer-review sessions within the research team resulted in the decision to exclude the metrics referring to:

- Utilisation of resources because the effects of resource scarcity regard the whole set of system objectives and not only safety. For example, a shortage of resources can lead to the reduction of production space or problems with service/product quality levels while safety levels are maintained. Therefore, since the research team's focus is on safety and the data collection about other system objectives were outside the scope of the project, the application of the particular metric was deemed as unfeasible. Nonetheless, the metric to be developed for the process level gaps between WaI and WaD, which comprises the PhD research conducted with the Delft University of Technology, focus on multiple system objectives and is expected to generate respective results through modelling and simulation tools.
- Complexity of socio-technical systems. The particular metric, although it is seen as more inclusive than the ones detected in the literature reviewed, necessitates further development and validation. Especially, the dynamic nature of multiple interactions and the formula element corresponding to human performance require more detailed research and further clarification (e.g., more detailed quantification of the parameters of the perceived complexity).

Regarding the metric of risk control effectiveness, a guidance document with example applications was prepared to support its implementation. The AVAC-SMS and AVAC-SCP metrics were converted into online questionnaires through the Qualtrics platform. Each questionnaire included an introductory text explaining the goal of the research, the anonymity of the participants, the voluntary nature of participation, and the expected benefits for the organisation. The right of each participant to withdraw the data after completing the surveys was not stated because the researchers did not record any identification information that would allow detecting the data set of a specific participant. However, since each company was given a unique code to participate in the surveys, we acknowledged to the contact persons that the withdrawal of the data was feasible only cumulatively for the whole company.

Two companies requested the translation of the online questionnaires to their local language: one company participating only in the AVAC-SMS survey and another one participating in the AVAC-SCP survey. The translations were performed by native, qualified persons and the translated questionnaires were finalised after an evaluation by the respective companies. The original versions of the questionnaires were pilot tested with the support of project partners and modified accordingly. The online versions of the questionnaires in English are available through the links shown in Appendix A.

4.2 Data collection, sample and processing

The two following sections explain the data collection, sample and processing regarding the AVAC-SMS and AVAC-SCP metrics. The metric referring to the risk control effectiveness was applied only by one company. Therefore, we were not able to analyse and compare results from the particular metric. In total 19 large and SME companies participated at least in one of the surveys: 14 from Europe, 2 from North America, 1 from Africa, 1 from Asia, and 1 from the Pacific Region. The company types, sizes and numbers participated in the surveys are presented in Table 2.

Company type	Size	AVAC-SMS	AVAC-SCP
Air Operator	Large	3	4
	SME	7	6
Air Navigation Service Provider	Large	1	1
Other (airports, ground handlers etc.)	Large	6	5

Company type	Size	AVAC-SMS	AVAC-SCP
	SME	1	0
Total	Large	10	10
	SME	8	6

Table 2: Sample distribution

4.2.1 Application of the AVAC-SMS

As explained in section 3.1 above, the questionnaires were related to three areas of SMS (Institutionalization, Capability, and Effectiveness). They were offered at three different resolution levels yielding a total of nine questionnaires with respective estimated completion times (Figure 2).; the latter were communicated to the companies to inform their decision-making about the resources they would invest in the SMS assessment. It is clarified that the Task level concerned, the indicated time of 4 hours reflects the duration of filling the questionnaire after the respondent has collected all relevant SMS documentation and logs (e.g., audit and training reports).

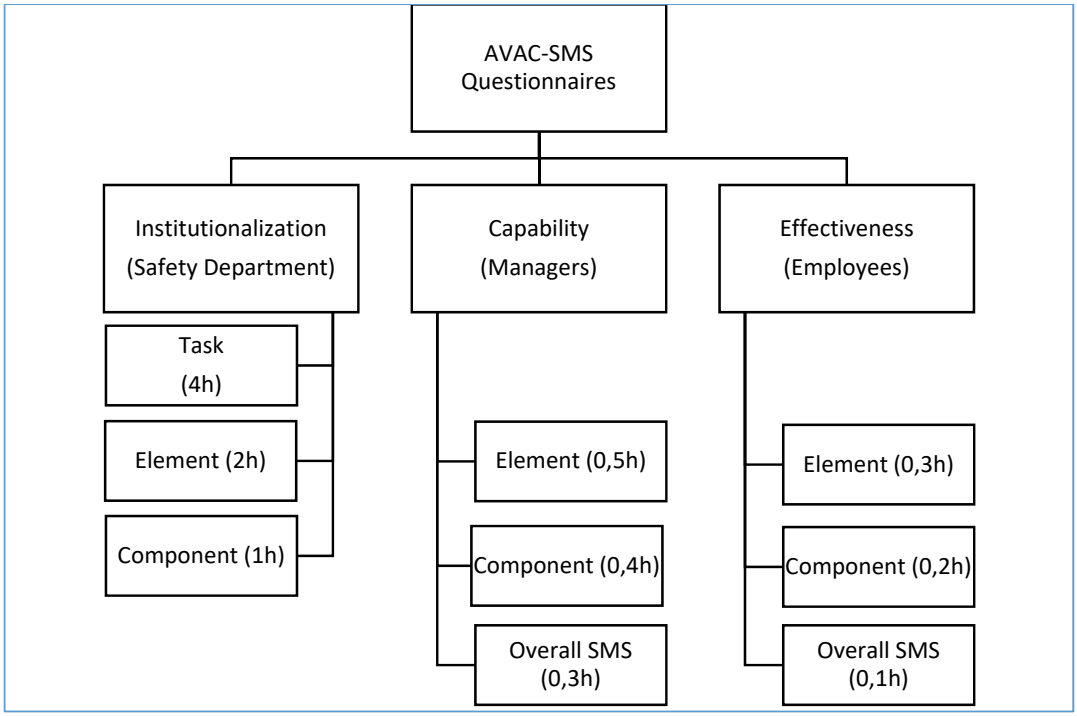

Figure 2. Overview of the AVAC-SMS questionnaires; completion time is reported in brackets

Table 3 presents the distribution of questions for the task level of the institutionalization dimension. The task level included compliance and implementation questions as well as time and process dependencies. The 149 questions (Table 4) were divided over three aspects: Design (i.e. compliance), Implementation (i.e. realization of design) and Dependencies (i.e. observing SMS process interfaces and timeliness). The different numbers of questions per SMS element are attributed to the various levels of description of the respective process in the Safety Management Manual (ICAO, 2013) and were finalised based on the comments received during the design of the metrics (see section 2 above).

SMS Institutionalization – Task Level			
SMS	Components	Elements	# Questions/Tasks
	Safety Policy & Objective (PO)	Management Commitment and Responsibility (MCR)	30
		Accountabilities and Responsibilities (AR)	8
		Resources and Key Personnel (RKP)	11
		Emergency Response Plan coordination (ERP)	13
		Documentation (SD)	11
		Subtotal	73
	Safety Risk Management (RM)	Hazard Identification (HI)	11
		Risk Assessment and Mitigation (RAM)	12
		Subtotal	23
	Safety Assurance (SA)	Performance Measurement and Monitoring (PMM)	20
		Change Management (CM)	10
		Continuous Improvement (CI)	6
		Subtotal	36
	Safety Promotion (PR)	Training and Education (TE)	8
		Communication (COM)	9
		Subtotal	17
	SMS	Total	149

Table 3. Overview of question numbers per component and elements.

Apart from the task level that was the one with the highest resolution, a fixed number of questions were presented for the Institutionalization at the Element and Component levels. In alignment with the dimensions assessed through the Task-level questionnaire, four questions were asked per element/component in correspondence with the four following dimensions:

- Design (i.e. according to standards)
- Implementation (i.e. realization of design)
- Timeliness (i.e. implementation activities at the proper time)
- Dependencies (i.e. use of inputs/outputs from other SMS elements/components)

Similarly, for Capability, there were six dimensions measured per element/component/overall SMS:

- Skills (i.e. staff knowledge and competencies to implement SMS tasks assigned)
- Means (i.e. availability of equipment and resources to implement SMS)
- Conflicts (i.e. different persons implementing SMS tasks but with divergent or opposite practices)
- Information (i.e. availability of information required to execute SMS tasks)
- Timeliness (i.e. timely reception of information necessary to perform SMS tasks)
- Disturbances (i.e. degree of other internal or external disturbances affecting negatively the execution of SMS tasks)

For the SMS Effectiveness assessment, there were three dimensions the employees were asked to evaluate:

- Quantity (i.e. sufficiency of SMS deliverables)
- Quality (i.e. quality of SMS deliverables)
- Timeliness (i.e. reception of SMS deliverables when proper/needed)

The companies were free to determine the level of assessment that best matched their structure, size and resource capacity and select who and how many employees filled out the questionnaires. Table 4 shows the participation (denoted by “X”) and data points in brackets per questionnaire and company. The Institutionalization excluded (i.e. the specific questionnaires were targeted only to the safety department and a single data point was the minimum required), the participation of employees in the rest of the SMS areas was not representative of the population of most of the companies. Therefore, the results for the whole sample could be only indicative.

Company Code	Institutionalization assessment level (sample size in brackets)			Capability assessment level (sample size in brackets)			Effectiveness assessment level (sample size in brackets)		
	Task	Element	Component	Element	Component	SMS	Element	Component	SMS
10629		X (3)	X (9)		X (4)	X (4)		X (1)	X (1)
10862		X (1)	X (1)		X (8)			X (4)	
12179		X (4)			X (2)			X (8)	
12821	X (2)	X (1)			X (5)			X (6)	
12903	X (2)	X (3)			X (32)			X (101)	
13567	X (1)				X (2)			X (15)	
15108	X (7)	X (4)	X (4)	X (28)			X (37)		
15521	X (1)	X (1)		X (3)	X (1)			X (1)	
15634	X (3)	X (3)	X (1)	X (54)	X (11)	X (13)	X (31)	X (28)	X (15)
16539	X (1)	X (1)	X (1)	X (2)	X (1)				X (53)
16652		X (1)	X (1)						
17029	X (3)			X (3)			X (2)		
17387	X (1)	X (1)	X (1)	X (5)	X (4)	X (4)	X (2)	X (2)	X (2)
19790	X (1)	X (1)							
20132		X (1)	X (1)	X (1)	X (1)		X (2)		
21381									X (1)
24113	X (1)				X (9)			X (5)	
24144	X (1)				X (2)			X (5)	

Table 4. Participation in each of the nine SMS questionnaires.

The institutionalization questionnaires were filled by the safety management department of each company which was requested to fill in at least two out of the three SMS assessment levels (i.e. task, element and component). The latter was to afford comparisons of the results yielded from different assessment levels and, possibly, allow companies to select a certain level of detail that would be most appropriate for their available resources. In general, the aim, on the one hand, was to check the consistency between different levels of assessment, and, on the other hand, to respect the resource and time limitations of the companies.

Regarding the other two SMS assessment areas, companies were invited to engage in the survey multiple managers (i.e. SMS capability) and work floor staff (i.e. SMS effectiveness). Companies were invited to fill out one capability and effectiveness questionnaire at any of the different SMS assessment levels out of the three available (i.e. element, component, overall SMS). As shown in Table 4, irrespective of the instructions provided, a few companies opted to fill in capability and effectiveness questionnaires at more than one levels, as with the institutionalization. Due to the limited sample, we were not able to compare the scores between different assessment levels for the capability and effectiveness areas.

Most of the questions could be answered by entering a percentage between 0 and 100 in increments of 20%. Only the Design questions of the Task level had a binary choice of 0% or 100% because they were referring to specific SMS items that, naturally, are present or not; for example, an SMS policy can exist or not and the answer could not take any intermediate value for partial compliance. As multiple employees per company performed the questionnaires, data were averaged by omitting null responses. The responses per employee were only included if at least 75% of the questions were answered. The calculations were performed as follows (see Appendix B for the detailed formulas):

- Questions for each element, component or overall SMS per entry were obtained by combining the averaged responses for the questions in that particular element, component or the overall SMS.
- For each SMS capability and effectiveness questionnaire, data were averaged over employee answers to come to a single value per question and company.
- Population scores were obtained by averaging over company scores.
- The results were also calculated per SMS area and dimension assessed.

Additionally, aggregated values to obtain results at higher levels (e.g., deriving results at a component level based on element scores) were obtained by averaging over questions related to the corresponding element, component, or overall SMS. We expected that there would be no significant differences amongst the final scores calculated at the SMS, component or element levels of aggregation for a single questionnaire. This was checked by applying the Cronbach's Alpha to determine the degree of agreement, where a value of "1" amongst the scores would represent a complete agreement (i.e. companies can use the score of any level

of aggregation) and a value of "0" would correspond to a complete disagreement (i.e. the level to which a score is aggregated reflects a different SMS assessment score).

To examine associations between the constructs assessed through the questionnaires, we applied the Pearson's correlation coefficient between SMS scores as follows:

- Questionnaires of the SMS institutionalization at different levels (i.e. Task, Element and Component levels). As explained above, this would indicate to what degree companies could confidently use questionnaires of various resolution levels to assess the particular SMS area. For example, if an assessment at the level of SMS component would be strongly correlated with the results from an assessment at the SMS element level, where the former has fewer questions compared to the latter, then companies could choose to use the SMS component questionnaire to save resources needed for the surveys to assess their SMS institutionalization.
- Questionnaires representing the three different SMS assessment areas of Institutionalization, Capability and Effectiveness. Particularly, we were interested in examining the relationships between the pairs of Institutionalization-Capability, Institutionalization-Effectiveness and Capability-Effectiveness as a means to indicate possible mutual dependencies. For these calculations, we considered the scores available per company regardless of the resolution level of assessment. If a company opted for multiple assessment levels, we used the score generated from the data of the most detailed level.

4.2.2 Application of the AVAC-SCP

Three questionnaires targeted to the following aspects of safety culture development:

- Organizational plans: whether the company has designed/documented each of the prerequisites
- Implementation: the extent to which the prerequisites are realized by the managers/supervisors across various organizational levels
- Perception: the degree to which frontline employees perceive the effects of managers' actions related to safety culture

The companies were asked to fill out the questionnaires on a self-assessment basis, and they were instructed to assess all three aspects; the estimated time investment was 4 hours for the Organizational plans (Safety Department), 0,3 hours for the Implementation (Managers) and 0,17 hours for the Perception questionnaires (Frontline Employees). The companies were asked to consider the time investment to engage as many managers and employees as possible. Table 5 shows the participation per company and questionnaire (denoted by "X") and reports in brackets the data points per case.

The Organisational plans excluded (i.e. the specific questionnaires were targeted only to the safety department, and a single data point was the minimum required), the participation of employees in the rest of the SCP areas was not representative of the population of most of the companies. Therefore, the results for the whole sample could be only indicative.

Company code	Organizational plans (sample size in brackets)	Implementation (sample size in brackets)	Perception (sample size in brackets)
10629	X (2)	X (5)	X (11)
10862	X (1)	X (8)	X (3)
12179	X (4)	X (1)	X (9)
12821	X (3)	X (5)	X (6)
12903	X (2)	X (16)	X (39)
15108	X (5)	X (18)	X (49)
15521	X (1)	X (2)	X (1)
15634	X (3)	X (50)	X (196)
16539	X (1)	X (1)	X (1)
16652	X (1)		
17029	X (2)	X (2)	
17387	X (1)	X (1)	X (5)
20132		X (1)	
24113	X (1)	X (4)	X (10)
24144	X (1)		X (10)
25226	X (2)	X (81)	

Table 5. Participation in the three SCP questionnaires.

The possible responses for the Organizational plans' questionnaire were Yes/Partially/No. The questions for the Implementation and Perception questionnaires were on a 5-point Likert scale. Two variants were possible depending on the question; these variants were coded to allow calculations (Table 6). Responses coded with "0" were treated as missing values. The responses were coded identically so the two scales could be combined. Since companies were invited to ask multiple employees to participate in the same questionnaire, a single score for each question was obtained by taking the median response over employees.

Variant 1	Strongly agree	Agree	Undecided	Disagree	Strongly disagree	
Variant 2	Always	Almost always	Sometimes	Almost never	Never	Not applicable
Code	5	4	3	2	1	0

Table 6. Likert scales for the Implementation and Perception questionnaires.

The questions of the three questionnaires were grouped into so-called *Sub-cultures* (see section 3.2 above). Table 7 illustrates the number of questions per different subculture and questionnaire.

Sub-culture	Organizational plans (54 questions)	Implementation (55 questions)	Perception (10 questions)
General prerequisites	18	19	4
Just culture	6	6	2
Flexible culture	5	5	1
Reporting culture	9	9	0
Informative culture	9	9	2
Learning culture	7	7	1

Table 7. Number of questions per sub-culture for the three AVAC-SCP questionnaires

The overall results were grouped by subculture and calculated as the medians of all responses to the respective questions per element. Associations between the Implementation and Perception scores as well amongst the sub-cultures per assessment aspect were assessed by using the Spearman's correlation coefficient. As the two questionnaires do not have the same number of items, only the overall scores were correlated. These scores were determined by taking the median of the responses of all questions of the corresponding questionnaire.

5. Results

5.1 AVAC-SMS results

5.1.1 Reliability tests and overall scores per company

The results from Cronbach's Alpha suggested that the scores at various level of aggregation (i.e. Task, Element and Component) were highly correlated, as it can be appreciated from Table 8. As such, only the overall SMS score per questionnaire was used for further calculations. The different scores per SMS area are presented in Table 9; the scores yielded per company at the highest resolution level, where applicable, are marked in bold. Kolmogorov-Smirnov tests showed that the data were normally distributed without statistically significant differences across the sample ($p>0,05$). The data suggest that Institutionalisation scores ranged from 0,59 to 0,97 (N=17, M=0,81, SD=0,12), Capability yielded scores between 0,54 and 0,86 (N=15, M=0,72, SD=0,09), and the Effectiveness scores ranged from 0,57 to 0,94 (N=16, M=0,75, SD=0,11). The detailed scores per assessment area, level and dimensions were communicated to the companies through individual reports.

	Institutionalization assessment level			**Capability assessment level**		**Effectiveness assessment level**	
	Task	Element	Component	Element	Component	Element	Component
Cronbach's Alpha	1,00	1,00	0,99	1,00	1,00	1,00	1,00

Table 8. Cronbach's Alpha values for scores aggregated at different SMS levels

Company Code	Institutionalization			Capability			Effectiveness		
	Task	Element	Component	Element	Component	SMS	Element	Component	SMS
10629		**0,73**	0,86		**0,71**	0,78		**0,80**	0,73
10862		**0,75**	0,78		**0,64**			**0,71**	
12179		**0,86**			**0,65**			**0,83**	
12821	**0,97**	0,90			**0,84**			**0,94**	
12903	**0,64**	0,68			**0,74**			**0,68**	
13567	**0,92**				**0,86**			**0,91**	
15108	**0,64**	0,55	0,65	**0,58**			**0,72**		
15521	**0,59**	0,83		**0,75**	0,63			**0,85**	
15634	**0,98**	1,00	0,94	**0,68**	0,63	0,64	**0,84**	0,81	0,83
16539	**0,84**	0,82	0,77	**0,80**	0,89				**0,68**
16652		**0,94**	0,95						
17029	**0,88**			**0,54**			**0,61**		
17387	**0,89**	0,83	0,85	**0,80**	0,86	0,86	**0,76**	0,81	0,90
19790	**0,89**	1,00							
20132		**0,83**	0,76	**0,80**	0,51		**0,57**		
21381									**0,8**
24113	**0,82**				**0,73**			**0,58**	
24144	**0,68**				**0,67**			**0,70**	
Average per column (calculated for N≥5)	0,81	0,82	0,82	0,71	0,72	N/A	0,70	0,78	0,79
Average per area (bold values)	0,81			0,72			0,75		

Table 9. SMS-level scores per company and questionnaire

5.1.2 Institutionalization

The results included in this section are presented graphically in Appendix C. At the component level of assessment (Figure C.1), the overall SMS score was 82,7% with Policy & Objectives (PO) and Safety Assurance yielding about 85% and Risk Management and Promotion (PR) scoring about 80% each. The dimensions concerned (Figure C.2), Design yielded the highest score (94,2%), followed by Implementation (84,4%), Timeliness (81,7%) and Dependencies (70,5%).

The findings from the assessment at the element level suggest that the picture regarding the differences across dimension scores remained the same (Figure C.4), and it provided a similar score for the overall SMS (82,7%). The picture per element (Figure C.3) revealed that Management Commitment and Responsibility, Resources & Key Personnel (RKP), Safety Documentation (SD), Hazard Identification (HI), Risk Assessment and Mitigation (RAM) and Training & Education (TE) were the ones with scores higher than the overall average, whereas the rest of the elements scored lower than the average. The elements with the two highest scores were HI (86,7%), and RAM (85,2%) and the ones with the lowest scores were Change Management (CM) (74,9%) and Continuous Improvement (CI) (79,0%).

At the highest resolution level of SMS tasks, the overall score was (83,9%) with almost equal percentages of the Design, Implementation and Dependencies dimension scores (Figure C.6). The elements which scored higher than or equal to the overall score (Figure C.5) were Emergency Response Planning (ERP), SD, HI and TE. The two lowest performed elements were Performance Measurement and Monitoring (PMM) (80,0%) and Communication (COM) (78,8%). When examining the dimensions per element (Figure C.6), the best-designed ones were SD, RAM and CM, whereas RKP yielded the lowest score. The implementation concerned, SD and TE scored visibly higher than the overall percentage and CM was rated lowest compared to the rest of the elements. Regarding the dependencies dimension, the highest scores were observed for RKP and ERP, and Communications had the lowest score. Appendix D reports the population results per SMS task; the top 25% of the scores are coloured in green and the lowest 25% of the scores in yellow.

5.1.3 Capability

The results included in this section are presented graphically in Appendix E. The overall SMS capability at the component level of resolution was 72,0% without major differences amongst the scores per component (Figure E.1). The dimensions concerned (Figure E.2), Skills and Means had the highest scores (81,5% and 78,0% respectively) whereas the Disturbances scored with 57,5%; it is clarified that the latter score reflects the extent to which disturbances do not affect the implementation of SMS activities. At the element level of resolution (Figure E.3), the overall capability score was calculated lower (70,7%); TE and PMM yielded the highest capability scores (77,8% and 75,9% correspondingly), and CM had the lowest capability percentage of 67,3% followed by Accountabilities & Responsibilities (AR), RKP and CI with scores around 68%. Regarding the dimensions (Figure E.4), their differences remained similar to the ones revealed by the assessment at the component level of resolution.

The least detailed assessment level concerned (Figure E.5), there were not enough data points to perform calculations. From a qualitative view of the respective graphs, it seems that the SMS capability scored higher than the element and component resolutions and, although the relative scores of Skills, Means and Disturbances remained similar to the scores obtained by the higher resolutions, the Information and Timeliness dimensions were rated as higher.

5.1.4 Effectiveness

The results included in this section are presented graphically in Appendix F. At the component assessment level (Figure F.1), the overall SMS effectiveness scored 78,2% with PO performing lowest (75,0%) and PR highest (81,4%) across the various components; the dimensions of quantity, quality and timeliness did not differ remarkably (Figure F.2). The element level concerned (Figure F.3), the overall SMS score was lower (69,8%) than the component resolution level, with the elements of HI and RAM yielding the highest scores (81,4% and 77,2% respectively); the lowest effectiveness was recorded for TE (64,5%), PMM (65,8%) and AR (64,9%). In this case, too, the dimensions did not show notable differences (Figure F.4). The lowest resolution assessment resulted in the score of 78,8% for the SMS and almost equal distribution of the values across the three dimensions (Figure F.5).

5.1.5 Statistical tests

The correlations between the pairs of the three resolution levels of the institutionalization assessment showed a high agreement: Task-Element (N=8, r=0.748, p=0.033), Element-Component (N=8, r=0.853, p=0.007). The Task-Component pair had only four data points and was not included in the calculations. The correlations between the three different constructs (i.e. Institutionalization, Capability, and Effectiveness) were not statistically significant.

5.2 AVAC-SCP results

The overall scores per company and assessment area are presented in Table 10, and the results per area and question are reported in Appendix G. In the particular Appendix, the top 25% fully or partially documented Organizational plans are coloured in green and the lowest 25% in yellow; a similar colour coding is used for the Implementation and Perception scores higher or lower than the median 4. It is noticed that the scores of negatively formulated questions have been inverted. The detailed findings per sub-culture and area assessed per company were communicated to the companies through individual reports.

Figures 3, 4 and 5 show the overall sample picture regarding Organisational Plans, Implementation and Perception respectively. The results suggest that Organisational Plans were about 83% fully or partially present (N=15, M=82,6, SD=12,83); the scores ranged from 53% to 100%. Just Culture prerequisites were the least represented at the level of 67%, and Reporting Culture prerequisites were 94% fully or partially included in the organisational plans. The rest of the subcultures concerned, organisational plans were fully or partially existent in 85% for General prerequisites, 80% for Flexible culture, 86% for Informative culture and 78% for Learning culture.

The Implementation yielded a median of 4 out of 5 in overall and across all subcultures. Employees Perceptions were in overall at the median level similar to Implementation, but staff rated Just Culture elements with the lowest score of 3.5 and the Flexible Culture elements with the highest score of 4.5.

Company Code	Organizational plans (%)		Implementation	Perception
	Yes	Partial	(median)	(median)
10629	87	0	4	4
10862	64	0	4	4
12179	85	1	4	4
12821	91	1	5	4
12903	53	5	4	4
15108	71	2	4	4
15521	68	0	4	4,5
15634	100	0	4	4
16539	88	1	4	4
16652	98	0		
17029	79	0	4	
17387	96	0	5	4
20132			4	
24113	77	3	4	4
24144	75	1		4
25226	93	0	4	

Table 10. SCP median scores per company and questionnaire

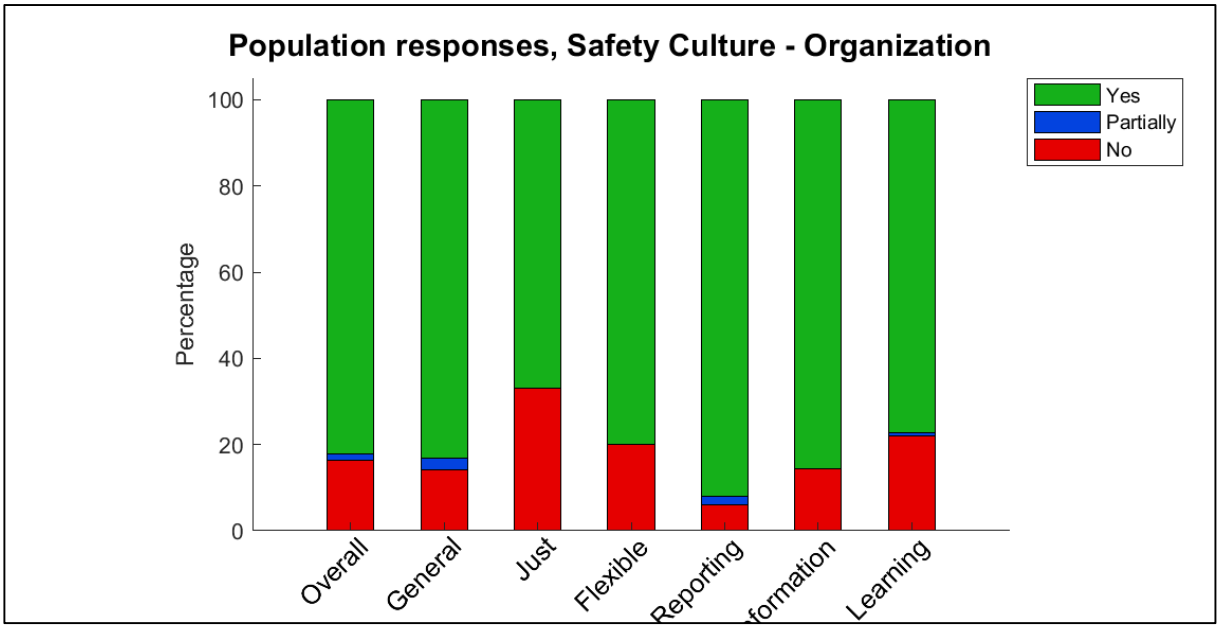

Figure 3. Population scores for the Organizational plans

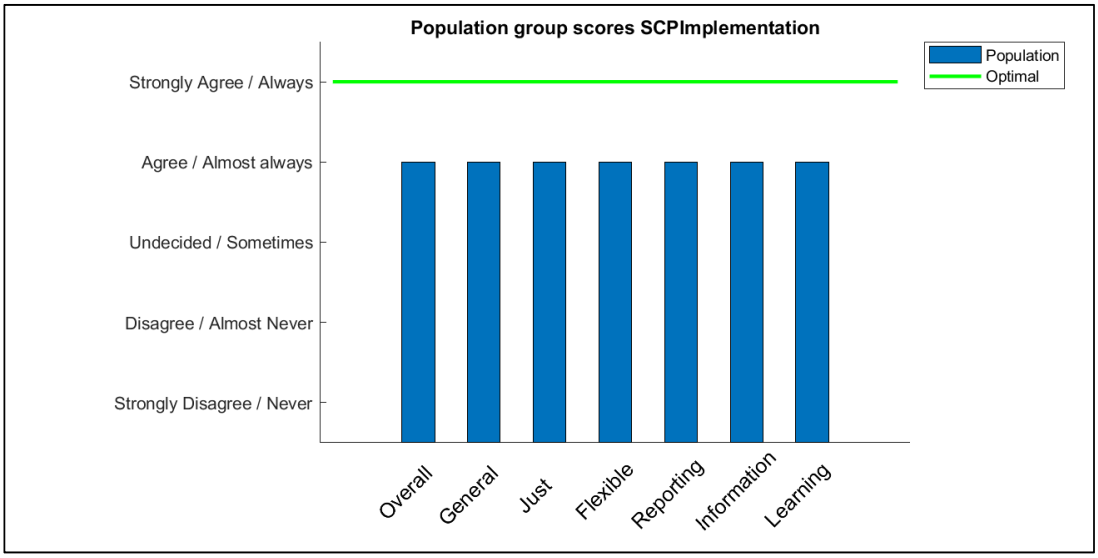

Figure 4. Population scores for the Implementation questionnaire.

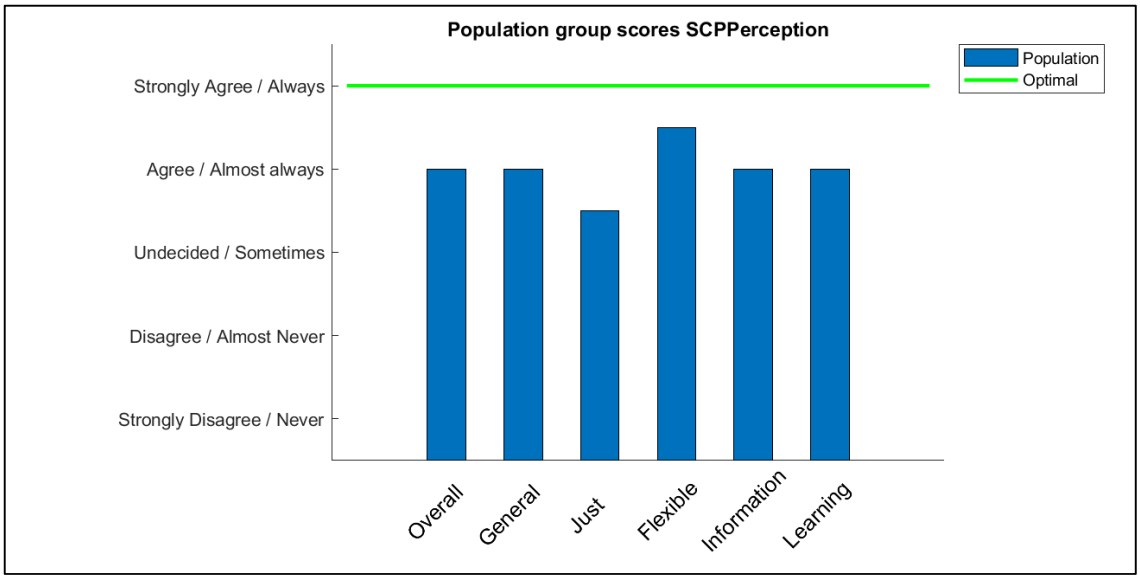

Figure 5. Population scores for the Perception questionnaire.

The Spearman's correlation coefficients between the Implementation and Perception scores in overall and per subculture per company were not statistically significant. The correlations amongst subcultures within the Implementation and Perception aspects resulted in non-significant results for the latter, whereas regarding Implementation, the following significances were detected:

- General prerequisites were found associated with Just culture (N=14, r=0,634, p=0,015) and Informative culture (N=14, r=0,534, p=0,049)
- Just culture was additionally associated with Reporting culture (N=14, r=0,599, p=0,024), Informative culture (N=14, r=0,885, p=0,000) and Learning culture (N=14, r=0,703, p=0,005)
- Learning culture was also found correlated with Reporting culture (N=14, r=0,637, p=0,014) and Informative culture (N=14, r=0,736, p=0,003)

6. Discussion

6.1 AVAC-SMS metric

Although the companies did not show statistically significant differences in their SMS scores across all three assessment areas, the sample averages showed a distance between the area of Institutionalization and the areas of Capability and Effectiveness. It must be noticed that the scores between these areas must be read as follows regarding the WaI-WaD gaps:

- The Institutionalization score (0,81) shows a 1-0,81=0,19 (or 19%) gap from the ideally designed and implemented system according to standards, briefly referred as ideal system hereafter. The ideal system assumes not only compliance but also effective implementation and added value of SMS to the organization.

- The Capability score (0,72) refers to the degree the existing Institutionalization activities can be fully realized (i.e. 72% level of realization). This means that the overall distance of the Capability (i.e. managers) from Institutionalization (safety department) is 1-0,72=0,28 (or 28%) and from the ideal system is 1-(0,72*0,81) =1-0,58=0,42 (i.e. 42%).

- The Effectiveness score (0,75) refers to the degree the employees perceive positively the SMS products that managers deliver (i.e. 75% value of the SMS products delivered). This means that the distance of Effectiveness (i.e. employees) from Capability (i.e. managers) is 1-0,75=0,25 (or 25%), from Institutionalization (i.e. safety department) is 1-(0,75*0,72) =1-0,54 =0,46 (or 46%), and from ideal system is: 1-(0,75*0,72*0,81) =1-0,44 = 0,56 (or 56%).

The final figure of the third bullet point above (i.e. 56%) can be roughly seen as the total SMS assessment score. However, this number can be only used for illustrative purposes and absolute measurement since it has not been internally or externally validated. The fact that there were no significant correlations amongst the Institutionalization, Capability and Effectiveness means that higher or lower performance of companies in one SMS area was not associated with the scores of the rest of the areas. This indicates that the three constructs are independent of each other and they measure different aspects.

When considering the more detailed results per area, the overall SMS institutionalization scores were comparable regardless of the level of assessment (i.e. Tasks, Components or Elements). However, the dimensions evaluated through the Component and Element level questionnaires revealed that Design (i.e. compliance to standards) scored considerably higher than the other dimensions and Dependencies (i.e. sharing and usage of deliverables generated by other SMS processes) collected the lowest rates. The Implementation and Timeliness scores fell in about the middle between Design and Dependencies. This suggests that companies adhere to planning their SMS elements and components as prescribed in the standards and they are close to its implementation as intended, but they might not have operated their SMS by adequately adopting a systems perspective that also considers the timeliness of activities and mutual dependencies. However, the Design, Implementation and Dependencies (i.e. time and input/output dependencies combined) did not differ when assessed at the most detailed level of SMS processes. This discrepancy might be attributed to the different types of questions posed to the participants; at the component and element assessment levels, the researchers used wording that was directly linked to the concepts of design, implementation, timeliness and dependencies, which might be perceived differently by various assessors.

Moreover, in addition to the gaps between the dimensions, there were differences amongst the SMS elements in overall and within each dimension. Although the results were not identical between the element and task levels of assessment, it is worth to notice that the former level concerned, Hazard Identification (HI) and Risk Assessment and Mitigation (RAM), which belong to the same SMS component and are seen by the industry as highly important, yielded the highest scores. HI was also found amongst the highest scoring elements in the task assessment level along with higher-than-average scores of Safety Documentation (SD) and Training & Education (TE) in both levels of assessment. Overall, the differences across and within dimensions, elements and components denote that companies, explicitly or implicitly, did not give the same

gravity to the various SMS items. Although there is no empirical research to support a weighing between SMS components and elements, perhaps the lack of resources in combination with different perceptions about the contribution of the various SMS processes to achieve safety objectives might have driven the investment of company efforts differently across SMS items and dimensions.

Regarding the SMS capability, the scores in overall and per parameter did not differ remarkably between the element and component levels of assessment. The fact that Skills and Means parameters had the highest scores indicates that companies focus much on the competencies of personnel and the equipment available to perform their SMS tasks. However, the low score of Disturbances means that managers were not always able to concentrate on their SMS activities due to external factors. When observing the differences amongst the scores of elements, it seems that management tasks were highly focused on TE and Performance Measurement & Monitoring (PMM), which rather reflect the overall emphasis of the industry on staff skills and measuring safety-related aspects. On the other hand, the softer SMS elements such as Continuous Improvement and Change Management scored lower; this possibly signals that managers preferred to steer their efforts and resources to elements with more immediate and visible results.

The SMS effectiveness concerned, the scores were similar for the component and overall SMS levels of assessment, but higher for the element level of assessment, which was the most detailed one. The component assessment level concerned, the fact that Policy & Objectives (PO) scored the lowest whereas Promotion (PR) scored the highest maybe reflects the different levels of employees' affection to the corresponding SMS activities. The former component regards mainly managerial tasks the deliverables of which might not be immediately or visibly available at the workforce to the same extent as safety communication, training and education activities. The latter naturally involve a higher degree of interaction with frontline employees. However, the results from the element level of assessment were considerably different; employees perceived the effectiveness of the Risk Management elements as the highest, while they rated as lowest the TE that belongs to the PR component. Although the researchers cannot explain these differences, it seems that especially for the effectiveness area of SMS, the level of assessment resolution affects the results dramatically. This might be a result of a different understanding across personnel of what each element and component entails; the questionnaires administered included brief descriptions of each SMS component and element, but this proved rather insufficient.

6.2 AVAC-SCP metric

The overall results regarding Organizational plans showed that companies had adequately included most of the Safety Culture Prerequisites (SCP) in their documentation. However, the fact that Just culture plans scored the lowest and Reporting culture plans were found with the highest percentage indicates that companies might have not completely recognized that an environment of fairness is a precondition for an effective reporting system. The SCP referring to the Reporting culture regard the characteristics of the reporting policy and system; however, these might not deliver the expected outcome if the Just culture prerequisites lag significantly behind. Also, the scores per company showed that there might be big differences between specific organisations; nonetheless, the effects of national and local cultures should not also be neglected.

The level of SCP implementation was the same high as the organizational plans and quite uniform across the companies and sub-cultures. Thus, the discrepancy between Just and Reporting cultures detected in SCP plans was not found in the implementation. Comparing the subculture differences detected in the SCP plans, it seems that managers, on the one hand, were implementing more just culture elements and fewer reporting culture elements than prescribed in the company documentation; the former can be seen as a positive gap, but the latter as a negative one.

Furthermore, although the perceptions were at the same overall level with implementation, it seems that employees perceived the organizational environment as less fair and more flexible than managers claimed. This finding indicates a gap that can have detrimental effects over time; managers who perceive themselves as advocates of just culture might not realize that workforce does not observe this in practice, and employees might believe that they have more room for flexibility than managers have offered.

The statistical associations concerned, the fact that the implementation of Just culture prerequisites was found significantly associated with all subcultures except Flexible subculture suggests the strong relationships between the fair working environment and the rest of the cultural aspects. Although the correlations detected do not mean causality, the particular finding seems supportive of the efforts placed recently to foster a just culture within organizations. The rest of significant associations concerned, the main message is rather that, in general, the different aspects of safety culture cannot be seen as completely independent from each other; this means, that a company focusing vastly on one subculture and underestimating the importance of other subcultures might not yield the maximum benefits.

7. Conclusions

The application of the AVAC-SMS and AVAC-SCP metrics showed that they have adequate sensitivity to capture any gaps between WaI and WaD amongst different organizational levels and across organizations. Also, the application of the metrics revealed interesting differences amongst the various aspects measured: Design, Implementation, Timeliness and Dependencies for the SMS and the subcultures for the Safety Culture Prerequisites. However, the relatively small sample of companies and restricted number of managers and employees participating per company render the findings only indicative and not conclusive. Also, this limitation did not allow to perform comparisons between large companies and SMEs as well as amongst companies with different operational activities (i.e. airlines, air navigation service providers, airports and ground services).

Although the study described in this report was exploratory and not explanatory, and the design of the research with different options of assessment resolution might have threatened the precision and comparability of the findings, we believe that the results presented above in combination with the ones communicated to the companies can trigger the latter to investigate further their weaker areas and foster their activities related to SMS and SCP. Therefore, the AVAC-SMS and AVAC-SCP metrics are deemed as useful to organizations that want to self-assess their SMS and SCP levels and proceed to comparisons amongst various functions and levels and/or over time.

On the side of practicality, the various assessment options offered for the AVAC-SMS can accommodate the resources each SME or large company can invest in the application of the metric. Although the statistical tests showed significant associations between the options for the Institutionalization at the overall SMS score, the differences observed between the three options (i.e. Tasks, Components and Elements) when considering the scores yielded per element, component and dimension indicate that the level of resolution chosen depends on what the organization wants to measure. If the overall SMS score is needed, then even the lowest level of resolution can be used. However, if a company seeks for a deeper and more valid assessment, it is advisable to use the most detailed assessment option that can afford. Regarding the AVAC-SMS areas of Capability and Effectiveness, the sample was not sufficient to perform statistical tests between different levels of assessment to suggest whether the various resolutions lead to similar scores. However, this will be considered in the future application of the metrics. The AVAC-SCP concerned, the specific metric does not offer multiple assessment options, but the number of questions, especially the ones targeted to frontline staff, are seen as manageable for the companies.

Finally, although the research team required from companies to share their figures of the activity and safety data (e.g., number of safety incidents, volume of flights) as a means to check associations of these with the scores of the metrics, the data collected was insufficient to perform statistical calculations. Therefore, at this stage, we could not determine whether the AVAC-SMS and AVAC-SCP have any predictive validity. The researchers plan to run a second round of surveys to apply the metrics and collect safety/activity data from more organizations, hence we anticipate that we will be able to test the metrics against safety performance and activity figures Nonetheless, irrespective of the possible associations of the metrics with safety outcomes, their application and findings communicated in this report are supportive of their usefulness, practicality and potential value for the companies that are interested in assessing their SMS and SCP, reveal gaps amongst the specific assessment areas per metric and get insights into their strong and weak points to improve further the way they manage safety.

Acknowledgements

The researchers would like to express their deep thanks to the following organisations and individuals for the expertise, time and efforts put in the current work:

- Design of New Safety Metrics (in alphabetically ascending order of last name): Bruggen Job (LVNL, NL), Christeas Stavros (Aegean Airlines, GR), Coetzee Gerhard (AirAstana, KZ), Davis Patricia (TAG Aviation, UK), de Giovanni John (United Airlines, US), de Kort Paul (SAFEmap Europe, NL), Dijkstra Arthur (KLM Ops, NL), Hamelijnck Regine (EASA, DE), Hencks Joel (AeroEX, CH), Hiltermann Ewout (KLM Cityhopper, NL), Hofinger Gesine (Team-HF, DE), Hol Richard (Transavia, NL), Karampelias Dimitrios (Athens International Airport, GR), Kartmale Keith (MUAC, NL), Mills Charlotte (New Zealand CAA, NZ), Nolan Conor (Aer Lingus, IE), Norris John (Av8jet, UK), Overijnder Mark (Jetsupport, NL), Panagopoulos Ilias (NATO Support and Procurement Agency, HU), Panagoulias Michail (Goldair Handling, GR), Plioutsias Anastasios (National Technical University of Athens, GR), Popp Christian (JetBlue, US), Richardson Neil (Baines Simmons, UK), Rietveld Dave (Helicentre, NL), Stanly Rufus (WaxWingAvionics, IN), Trapp Duncan (CHC Helicopter, CA), van de Groep Andre (LVNL, NL), van Dooren Nico (Schiphol Airport, NL), van Hilten Gert Jan (KLM Ground Services, NL), van Kleef Eric (Van Kleef Consultancy, NL), van Maurik Ruud (MOD/MLA, NL), van Vliet Daniel (SkyServiceNetherlands, NL), vd Westhuizen Jaco (ATNS, ZA), Vergouwen Frank (Hill's Pet Nutition, NL).
- Application of New Safety Metrics (in alphabetically ascending order of company name): Aer Lingus (IE), Aeroporti di Roma (IT), Air Dolomiti (IT), ATNS (ZA), Av8jet (UK), Aviation Fuel Services-OFC (GR), Goldair Handling (GR), Kroger Co (US), TAG Aviation (UK), Napoli airport-GESAC (IT), Oceania Aviation (NZ), Orio al Serio airport-SACBO (IT), Schiphol Airport (NL), Shell aircraft international (NL), Star Wings (DE), THAI Airways (TA), United Airlines (US) and Volkswagen Airservice (DE).
- Review of this report (in alphabetically ascending order of last name): Robert J. de Boer (Northumbria University – Amsterdam Campus, NL), Gesine Hofinger (Team-HF, DE), Julius Keller (Purdue University, US), Charlotte Mills (Civil Aviation Authority, NZ), Selma Piric (HUFAG, NL), Anastasios Plioutsias (National Technical University of Athens, GR), Ruud van Maurik (MOD/MLA, NL), Alexei Sharpanskykh (TU Delft, NL).

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.

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

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

Fitzgerald, K., Miles, R., Porter, D., Scanlon, M. (2011). Key performance indicators for human factors in major hazard industries. IChemE Symposium Series, 156, 396-403.

Hollnagel, E. (2017). Contextual Control Model (COCOM). Retrieved 12 December 2017 from http://erikhollnagel.com/onewebmedia/COCOM.pdf.

Karanikas, N., Kaspers, S., Roelen, A., Piric, S., van Aalst R. & de Boer, R. J. (2017b). Five Novel Metrics to Support Safety Improvements in Current and Future Air Transport. AEROjournal, 2, pp. 21-27.

Karanikas, N., Kaspers, S., Roelen, A.L.C., Piric, S., & de Boer, R. J. (2016b). Review of Existing Aviation Safety Metrics, RAAK PRO Project: Measuring Safety in Aviation, Project S10931, Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands. DOI: 10.13140/RG.2.2.35967.41128

Karanikas, N., Kaspers, S., Roelen, A.L.C., Piric, S., van Aalst, R. & de Boer, R. J. (2016a). Results from Surveys about Existing Aviation Safety Metrics, RAAK PRO Project: Measuring Safety in Aviation, Project S10931, Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands. DOI: 10.13140/RG.2.2.22545.63847

Karanikas, N., Kaspers, S., Roelen, A.L.C., Piric, S., van Aalst, R. & de Boer, R. J. (2017a). Concept for the Design of New Metrics, RAAK PRO Project: Measuring Safety in Aviation, Project S10931, Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands. DOI: 10.13140/RG.2.2.26107.72489

Karanikas, N., Selma, P., de Boer, R. J., Roelen, A., Kaspers, S. and van Aalst, R. (2018). The AVAC-SMS Metric for the Self-assessment of Maturity of Aviation Safety Management Systems, Proceedings of the International Cross-industry Safety Conference, 2-3 November 2017, Amsterdam University of Applied Sciences, AUP Advances, 1(1), pp. 40-57, DOI: 10.5117/ADV2018.1.003.KARA

Karanikas, N., Soltani, P., de Boer R. J. & Roelen A.L.C. (2016c). Safety Culture Development: the Gap Between Industry Guidelines and Literature, and the Differences Amongst Industry Sectors, in Arezes, P. (ed.), Advances in Safety Management and Human Factors, Proceedings of the AHFE 2016 International Conference on Safety Management and Human Factors, July 27-31, 2016, Walt Disney World®, Florida, USA, Springer. pp. 53-63. DOI: 10.1007/978-3-319-41929-9_7

Kaspers, S., Karanikas, N., Roelen, A., Piric, S. & de Boer, Robert J. (in press). How Does Aviation Industry Measure Safety Performance? Current Practice and Limitations. International Journal of Aviation Management.

Kaspers, S., Karanikas, N., Roelen, A., Piric, S., van Aalst, R. & de Boer, R. J. (2016). 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), pp. 18-29. DOI: 10.5296/jss.v2i2.10437

Kaspers, S., Karanikas, N., Roelen, A., Piric, S., van Aalst, R. & de Boer, R.J. (2017). Measuring Safety in Aviation: Empirical Results about the Relation between Safety Outcomes and Safety Management System Processes, Operational Activities and Demographic Data, PESARO 2017: The Seventh International Conference on Performance, Safety and Robustness in Complex Systems and Applications, IARIA, pp. 9-16.

Körvers, P.M.W. (2004). Accident precursors: pro-active identification of safety risks in the chemical process industry. (PhD Thesis). Eindhoven University of Technology.

Muns, R. (2017). Indicators based on the effectiveness of risk control measures. (Unpublished BSc graduation thesis). The Netherlands: Amsterdam University of Applied Sciences.

NLR (2016). ASC-IT: Seven steps to improve your safety culture. NLR-CR-2016-228

OED. (2017). Oxford English Dictionary online. Retrieved 14 November 2017 from: http://www.oed.com/

Piric, S., Karanikas, N., De Boer, R. J., Roelen, A., Kaspers, S. & Van Aalst, R. (2018). How much do Organizations Plan for a Positive Safety Culture? Introducing the Aviation Academy Safety Culture Prerequisites (AVAC-SCP) Tool, Proceedings of the International Cross-industry Safety Conference, 2-3 November 2017, Amsterdam University of Applied Sciences, AUP Advances, 1(1), pp. 118-129, DOI: 10.5117/ADV2018.1.008.PIRI

Rasmussen, J., Svedung, I. (2000). Proactive risk management in a dynamic society. Karlstad, Sweden: Swedish Rescue Services Agency.

Reason, J. (1998). Achieving a safe culture: theory and practice. Work Stress 12(3), 293–306

Roelen, A., Van Aalst, R., Karanikas, N., Kaspers, S., Piric, S. & De Boer, R. J. (2018a). Effectiveness of Risk Controls as Indicator of Safety Performance, Proceedings of the International Cross-industry Safety Conference, 2-3 November 2017, Amsterdam University of Applied Sciences, AUP Advances, 1(1), pp. 175-189, DOI: 10.5117/ADV2018.1.012.ROEL

Roelen, A., Van Aalst, R., Karanikas, N., Kaspers, S., Piric, S. & De Boer, R. J. (2018b). Safety Metrics Based on Utilisation of Resources, Proceedings of the International Cross-industry Safety Conference, 2-3 November 2017, Amsterdam University of Applied Sciences, AUP Advances, 1(1), pp. 88-98, DOI: 10.5117/ADV2018.1.006.ROEL

Shaw, J. D., Delery, J. E., Jenkins, G. D., Gupta, N. (1998). An organizational-level analysis of voluntary and involuntary turnover. Academy of Management Journal, 41, 511–525.

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

Van Aalst, R., Karanikas, N., De Boer, R. J., Roelen, A., Kaspers, S. & Piric, S. (2018). Complexity of Socio-Technical Systems: concept for a uniform metric, Proceedings of the International Cross-industry Safety Conference, 2-3 November 2017, Amsterdam University of Applied Sciences, AUP Advances, 1(1), pp. 142-153, DOI: 10.5117/ADV2018.1.010.VAN

Appendix A

(The trial code to access all questionnaires is 99747)

AVAC-SMS metric

Institutionalization

Task level: https://hva.eu.qualtrics.com/jfe/form/SV_8bQSj9qW9FPsZyB

Element level: https://hva.eu.qualtrics.com/jfe/form/SV_bCynLcbVRcLCJRX

Component level: https://hva.eu.qualtrics.com/jfe/form/SV_cXQJNCm5yUlTDvf

Capability & Effectiveness

Element level: https://hva.eu.qualtrics.com/jfe/form/SV_eEyLfLqTZb2mrTT

Component level: https://hva.eu.qualtrics.com/jfe/form/SV_bOVTdEx8rpxS4Ml

Overall SMS: https://hva.eu.qualtrics.com/jfe/form/SV_3qLcituguPC0Tml

AVAC-SCP metric

Organisational plans: https://hva.eu.qualtrics.com/jfe/form/SV_4MGQaP63l3pZ2YJ

Implementation: https://hva.eu.qualtrics.com/jfe/form/SV_eM15QgtNqgS9Ng9

Perception: https://hva.eu.qualtrics.com/jfe/form/SV_82I5QEQFXBwde9D

Appendix B

All calculations are based on scores (the responses to the questions), the maximum score, the ratio, and distance from the maximum score. There were 149 questions each relating to a specific task. The response to task *i* is given by $resp_{t_i}$. The number of tasks in element *i* is given by nt_{e_i}, the number of elements in component *i* is given by ne_{c_i}, and the number of questions for each element or component for questionnaires are those level is denoted by n_q.

Each outcome measure is determined by first calculating a maximum score and a distance score, representing the distance between the scores and the maximum. The outcome measure O is then calculated as $O = 1 - \frac{D}{M}$. The measure can in some cases be weighted or unweighted, denoted by a superscript W or U, respectively. If these are equal then this will be omitted.

Institutionalization – Task level

Intermediate calculations

- Task

$$s_{t_i} = resp_{t_i}$$
$$m_{t_i} = 100$$

- Element

$$s_{e_i} = \sum_{t_j \in e_i} resp_{t_j}$$
$$m_{e_i} = 100 \cdot nt_{e_i}$$
$$r_{e_i} = \frac{s_{e_i}}{m_{e_i}}$$
$$d_{e_i} = m_{e_i} - s_{e_i}$$

- Component

$$s_{c_i} = \sum_{e_j \in c_i} s_{e_j}$$
$$m_{c_i} = \sum_{e_j \in c_i} m_{e_j}$$
$$r_{c_i} = \frac{s_{c_i}}{m_{c_i}}$$
$$d_{c_i} = m_{c_i} - s_{c_i}$$

- Overall SMS

$$s_{sms} = \sum s_{c_i}$$
$$m_{sms} = \sum c_i$$
$$r_{sms} = \frac{s_{sms}}{m_{sms}}$$
$$d_{sms} = m_{sms} - s_{sms}$$

SMS Scores

- Task

$$D = \sqrt{\sum_{i=1}^{n_t} \left(100 - resp_{t_i}\right)^2}, M = 100\sqrt{n_t}$$

- Element

$$D^U = \sqrt{\sum_{i=1}^{n_e} \left(100 - r_{e_i}\right)^2}, M^U = \sqrt{n_e}$$

$$D^W = \sqrt{\sum_{i=1}^{n_e} (d_{e_i})^2}, M^W = \sqrt{\sum_{i=1}^{n_e} (m_{e_i})^2}$$

- Component

$$D^U = \sqrt{\sum_{i=1}^{n_c} (100 - r_{c_i})^2}, M^U = \sqrt{n_c}$$

$$D^W = \sqrt{\sum_{i=1}^{n_c} (d_{c_i})^2}, M^W = \sqrt{\sum_{i=1}^{n_e} (m_{c_i})^2}$$

- Overall SMS

$$D_{sms} = 1 - r_{sms},\ M_{sms} = 1$$

Institutionalization, Capability, and Effectiveness – Element level

Intermediate calculations

- Element

$$s_{e_i} = \sum_{q_j \in e_i} resp_{q_j}$$
$$m_{e_i} = 100 \cdot nq_{e_i}$$
$$r_{e_i} = \frac{s_{e_i}}{m_{e_i}}$$
$$d_{e_i} = m_{e_i} - s_{e_i}$$

- Component

$$s_{c_i} = \sum_{e_j \in c_i} s_{e_j}$$
$$m_{c_i} = \sum_{e_j \in c_i} m_{e_j}$$
$$r_{c_i} = \frac{1}{ne_{c_i}} \sum_{e_j \in c_i} r_{e_j}$$
$$d_{c_i} = m_{c_i} - s_{c_i}$$

- Overall SMS

$$s_{sms} = \sum s_{c_i}$$
$$m_{sms} = \sum c_i$$
$$r_{sms} = \frac{1}{ne} \sum r_{e_i}$$
$$d_{sms} = m_{sms} - s_{sms}$$

SMS Scores

- Element

$$D = \sqrt{\sum_{i=1}^{n_e} \left(100 - r_{e_i}\right)^2}, M = \sqrt{n_e}$$

- Component

$$D^U = \sqrt{\sum_{i=1}^{n_c} \left(100 - r_{c_i}\right)^2}, M^U = \sqrt{n_c}$$

$$D^W = \sqrt{\sum_{i=1}^{n_c} \left(d_{c_i}\right)^2}, M^W = \sqrt{\sum_{i=1}^{n_e} \left(ne_{c_i} \cdot n_q \cdot 100\right)^2}$$

- Overall SMS

$$D_{sms} = 1 - r_{sms},\ M_{sms} = 1$$

Institutionalization, Capability, and Effectiveness – Component level

Intermediate calculations

- Component

$$s_{c_i} = \sum_{q_j \in c_i} resp_{q_j}$$
$$m_{c_i} = \sum_{e_j \in c_i} m_{e_j}$$
$$r_{c_i} = \frac{1}{ne_{c_i}} \sum_{e_j \in c_i} r_{e_j}$$
$$d_{c_i} = m_{c_i} - s_{c_i}$$

- Overall SMS

$$s_{sms} = \sum s_{c_i}$$
$$m_{sms} = \sum c_i$$
$$r_{sms} = \frac{1}{n_c} \sum r_{c_i}$$
$$d_{sms} = m_{sms} - s_{sms}$$

SMS Scores

- Component

$$D = \sqrt{\sum_{i=1}^{n_c} \left(100 - r_{c_i}\right)^2}, M = \sqrt{n_c}$$

- Overall SMS

$$D_{sms} = 1 - r_{sms},\ M_{sms} = 1$$

Capability, and Effectiveness – Overall SMS level

Intermediate calculations

- Overall SMS

$$s_{sms} = \sum resp_i$$
$$m_{sms} = \sum 100 \cdot n_c$$
$$r_{sms} = \frac{s_{sms}}{m_{sms}}$$
$$d_{sms} = m_{sms} - s_{sms}$$

SMS Score

- Overall SMS

$$D_{sms} = 1 - r_{sms},\ M_{sms} = 1$$

Appendix C

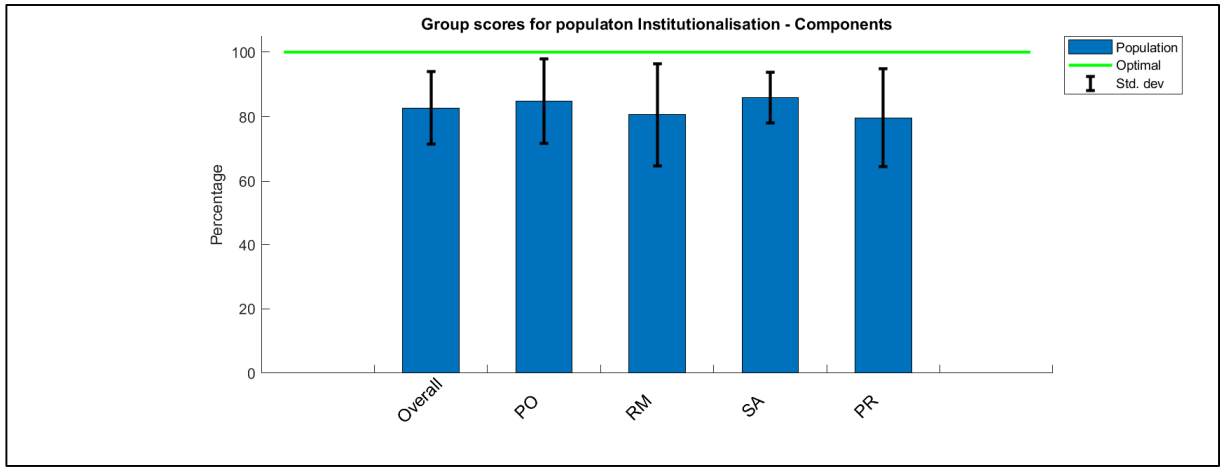

Figure C.1. Population results for each component of the Institutionalization questionnaire at the component level.

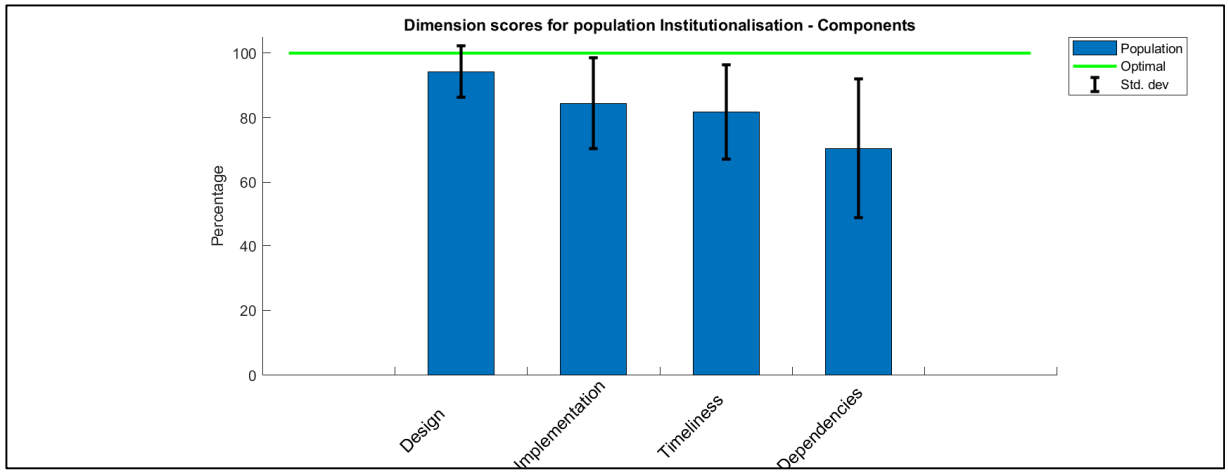

Figure C.2. Population results for each dimension of the Institutionalization questionnaire at the component level.

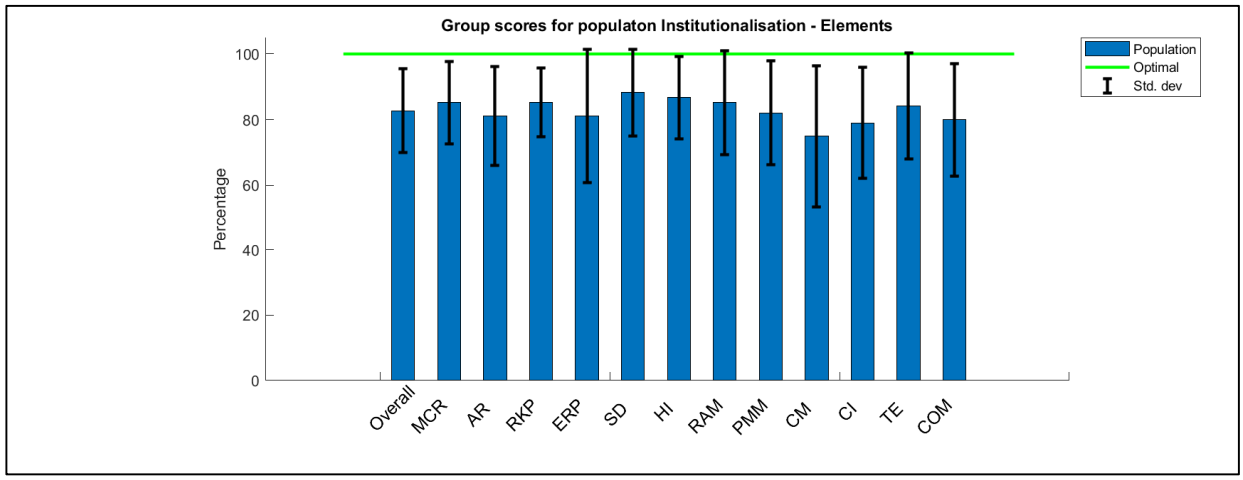

Figure C.3. Population results for each element of the Institutionalization questionnaire at the element level.

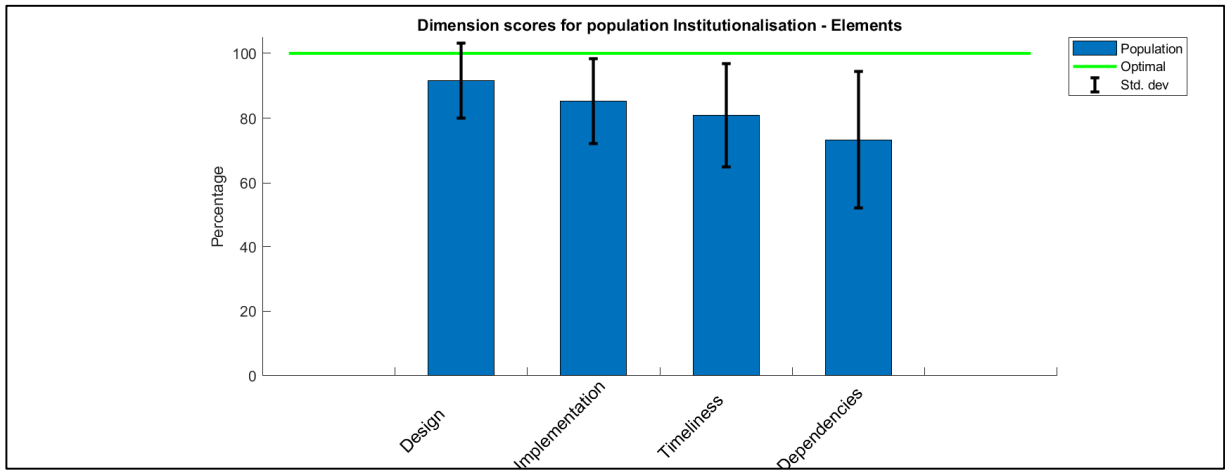

Figure C.4. Population results for each dimension of the Institutionalization questionnaire at the element level.

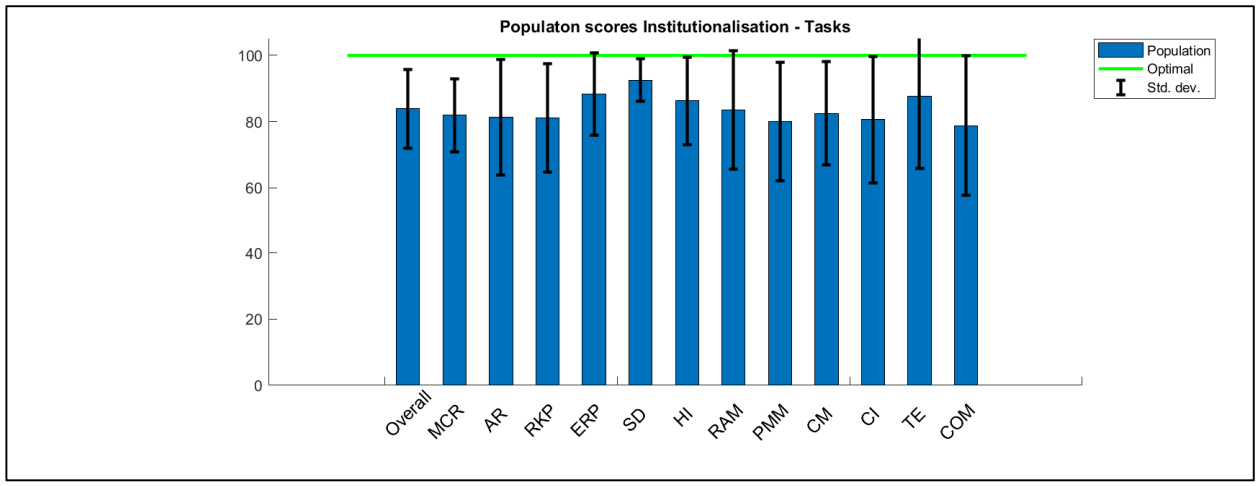

Figure C.5. Population results of the Institutionalization questionnaire at the tasks level grouped by element.

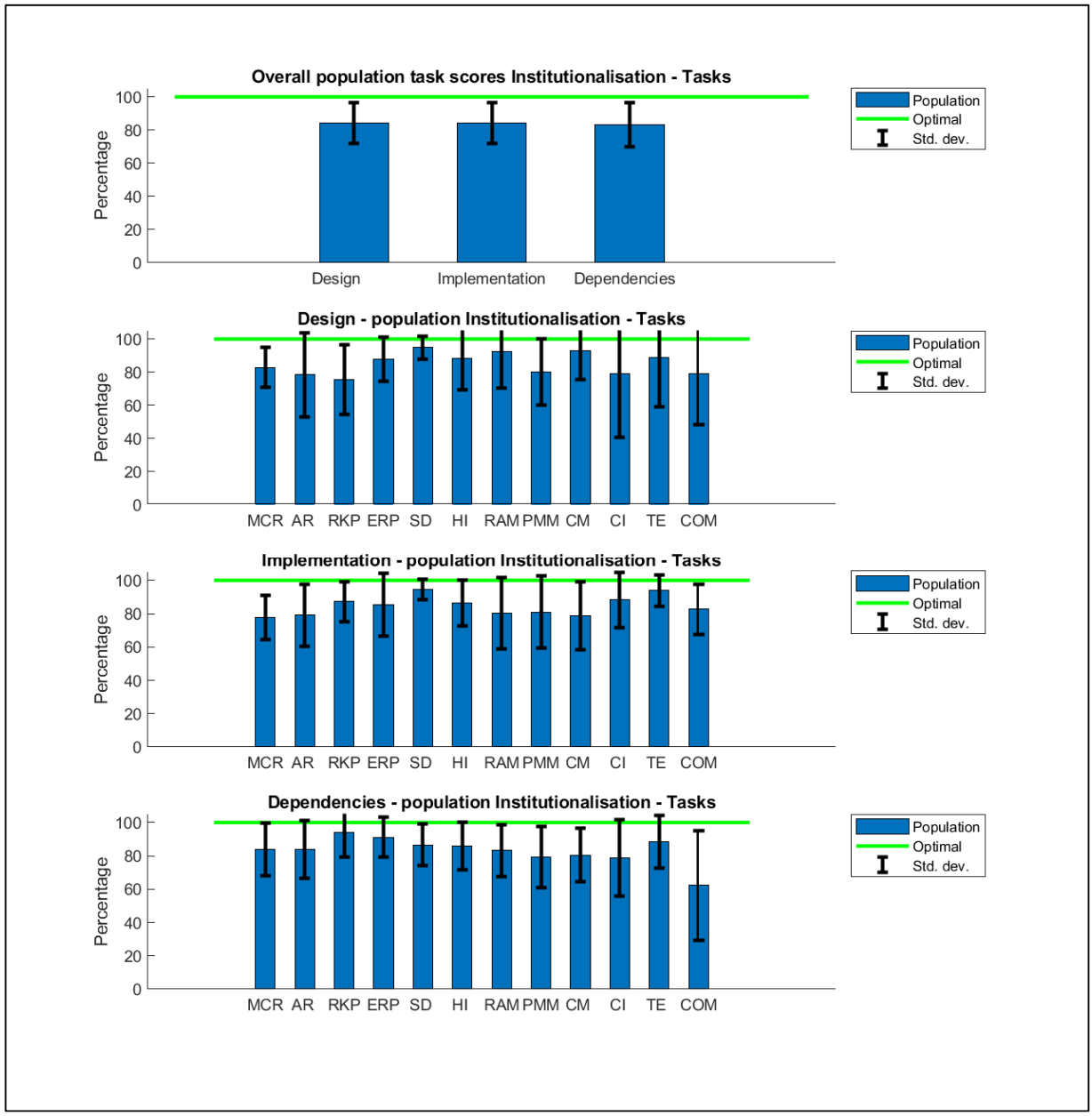

Figure C.6. Population results for the Institutionalization questionnaire at tasks level grouped by aspect (top panel) and each aspect across elements (bottom three panels).

Appendix D

Component	Element	Question Code	Question Text (Task)	Population Score [%]
Safety Policy & Objectives	Management Commitment & Responsibility	MCR01	There is a safety policy	100
		MCR02	The overall organisational policy views safety as one of the core business functions	95
		MCR03	Safety staff and officers participate in all planning and review management meetings (across all organizational levels and sections, as applicable)	73
		MCR04	Safety is a parameter in decision-making during all planning and review management meetings (across all organizational levels and sections, as applicable)	72
		MCR05	The possible need to change the safety policy has been always discussed by management during significant changes within the organization	75
		MCR06	Current safety policy is included in all safety education/training programs	91
		MCR07	There is a just culture policy	88
		MCR08	Types of unacceptable behavior have been defined in the just culture policy	100
		MCR09	Possible consequences of each type of unacceptable behavior have been defined in the just culture policy, or any other document connected to this policy	73
		MCR10	A right to appeal is part of the just culture policy or other document linked to this policy	65
		MCR11	Just culture policy has been established based on the maximum possible agreement from all organizational levels and functions (e.g., survey, workshops)	78
		MCR12	When applying the just culture policy, decision-making and argumentation are always and fully documented	67
		MCR13	The just culture policy is always applied based on combination and consideration of all available hard evidence and personal accounts	81
		MCR14	The just culture policy has been always implemented when unacceptable behavior was identified through observations, supervisory tasks, reports etc.	77
		MCR15	The just culture policy is included in all safety training/education programs	89
		MCR16	Safety objectives have been defined	98
		MCR17	All safety objectives are aligned with the safety policy	90
		MCR18	All new/changed safety objectives have been always balanced with other business objectives (e.g., efficiency, productivity)	81
		MCR19	There are policy/procedures for mandatory reporting	100
		MCR20	There are policy/procedures for voluntary reporting	91
		MCR21	Voluntary reporting is designed as confidential	61
		MCR22	There are policy/procedures for timeliness of feedback provision to staff who report voluntarily	55
		MCR23	There are policy/procedures for content and type of feedback provision to staff who report voluntarily (e.g, risk level assessed, planned actions)	59
		MCR24	Identification data of all staff submitting voluntary reports are always treated as confidential	88
		MCR25	Feedback to all staff submitting voluntary reports is provided	78
		MCR26	Content of all feedback given to staff submitting voluntary reports is according to respective policy/procedures	84
		MCR27	Feedback to staff submitting voluntary reports is provided within the defined timeframe	71
		MCR28	The policy for voluntary reporting refers to the the just culture policy	64
		MCR29	The voluntary reporting policy and procedures are included in all safety education/training programs	85
		MCR30	The mandatory reporting policy and procedures are included in all safety education/training programs	89
	Accountabilities & Responsibilities	AR1	Safety accountabilities are included in all job descriptions	78
		AR2	All safety accountabilities are realised (e.g., reporting to seniors the progress of safety improvements)	80
		AR3	All safety accountabilities correspond to the hierarchical structure of the organization (i.e. junior levels are accountable to senior ones)	89

		AR4	The safety accountabilities per function/role are part of all safety education/training programs	84
		AR5	Safety responsibilities are included in all job descriptions	76
		AR6	All safety responsibilities are realised (e.g., initiation and monitoring of safety improvements, enforcement and application of agreed safety rules)	74
		AR7	All safety responsibilities of staff are aligned with their authority to allocate resources, change tasks etc	75
		AR8	The safety responsibilities per function/role are part of all safety education/training programs	84
	Assignment of Resources & Appointment of Key Personnel	RKP01	The number of safety personnel required to implement and maintain the SMS I (e.g., safety officers, advisers, investigators) are defined (per organisational level or section, as applicable)	80
		RKP02	The competencies of safety personnel of every role are defined (e.g., safety manager, officers, advisers, investigators)	91
		RKP03	All foreseen safety personnel have been appointed	93
		RKP04	All currently appointed safety personnel fulfil all predefined competencies for their roles	91
		RKP05	Amount and type of technical equipment are defined for the implementation of safety policy (per organisational level or section, as applicable)	76
		RKP06	Specifications are defined for all technical equipment used for the implementation of safety policy (per organisational level or section, as applicable)	68
		RKP07	All foreseen technical equipment for the implementation of safety policy are available	79
		RKP08	All available technical equipment for the implementation of safety policy meet the predefined specifications	79
		RKP09	A budget for the implementation of safety policy and achievement of safety objectives is defined	53
		RKP10	The budget required for the implementation of safety policy and achievement of safety objectives is available and spent	97
		RKP11	Concurrent business needs have been always assessed when resources for the implementation of safety policy and achievement of safety objectives have been assigned/modified	93
	Coordination of Emergency Response Planning	ERP01	There is an Emergency Response Plan (ERP)	100
		ERP02	The ERP refers to interfaces between the organisational levels, departments, functions etc. involved	98
		ERP03	The ERP reflects/considers all possibly relevant internal organisational interfaces	92
		ERP04	The ERP defines the resources required internally per department, function etc. involved	82
		ERP05	All internal resources which are planned in the ERP are available (e.g., amount of staff, equipment, tools)	93
		ERP06	The ERP refers to interfaces with the external agencies, organisations, authorities etc. involved	98
		ERP07	The ERP reflects/considers all possibly relevant external interfaces	91
		ERP08	The ERP defines the resources required per external agent involved	62
		ERP09	All external resources included in the ERP are periodically verified	80
		ERP10	The periodicity for testing the ERP is defined	83
		ERP11	The ERP is always tested within the predetermined periodicity	82
		ERP12	The ERP considers all major levels and types of operational activities (e.g., day or night activities and operations, types of emergencies)	94
		ERP13	The ERP is included in all safety education/training programs of the internal sections, departments etc. involved	87
	SMS Documentation	SD01	There is SMS documentation	100
		SD02	The SMS documentation describes all SMS components and elements required by standards (i.e. what are the aims, objectives etc.)	95
		SD03	The SMS documentation includes or refers to methods/tools/procedures for the operationalisation of	93

			all SMS components and elements (i.e. how to implement)	
		SD04	The SMS documentation requires, or refers to, records/logs of all SMS activities	92
		SD05	All versions of SMS documentation have been approved by the senior management	97
		SD06	The SMS documentation has been updated with all latest changes required by standards, legislation etc.	95
		SD07	The newest version of the SMS documentation is accessible to workplaces of all staff	94
		SD08	All major SMS activities are recorded/logged	92
		SD09	All changes in SMS processes and activities have been always included in safety communication to the persons affected	84
		SD10	The SMS documentation scope, structure, contents etc. are included in all safety education/training programmes	90
		SD11	The SMS documentation has been always updated before effective dates of the SMS or externally imposed changes (e.g., legislation, standards) were due	82
Safety Risk Management	Hazard Identification	HI01	Internal sources for proactive hazard collection/identification have been defined	92
		HI02	Internal sources for reactive hazard collection/identification have been defined	98
		HI03	External sources for proactive hazard collection/identification have been defined	77
		HI04	External sources for reactive hazard collection/identification have been defined	82
		HI05	Risk registry shows that proactive hazard information is collected/identified from all predefined internal sources	87
		HI06	Risk registry shows that reactive hazard information is collected/identified from all predefined internal sources	89
		HI07	Risk registry shows that proactive hazard information is collected/identified from all predefined external sources	80
		HI08	Risk registry shows that reactive hazard information is collected/identified from all predefined external sources	83
		HI09	All information/data from all voluntary reports are used in hazard identification	85
		HI10	The proactive hazard identification procedures/policies are included in all safety education/training programs	86
		HI11	The reactive hazard identification procedures/policies are included in all safety education/training programs	82
	Risk assessment and mitigation	RAM01	A method for risk assessment has been defined	94
		RAM02	A method/procedure for the development of risk controls has been defined	89
		RAM03	The risk registry shows that risk assessment is applied for all safety hazards identified or changed	86
		RAM04	All risks are always assessed according to the predefined methods/procedures	87
		RAM05	All risk controls are always developed according to the predefined methods/procedures	86
		RAM06	Deadlines are always defined for the implementation of all risk controls	75
		RAM07	Risk controls are always implemented within the timeframe decided	68
		RAM08	The criteria for risk acceptance are customized to the authority of all job functions/roles	76
		RAM09	Risk controls are always developed or changed when risk assessments conclude to non-acceptable risks	87
		RAM10	All procedural and technical risk controls have been included in working procedures/temporary notices	78
		RAM11	The risk assessment policies/procedures are included in all relevant safety education/training programs	86
		RAM12	The policies/procedures for the development of risk controls are included in all relevant safety education/training programs	84
Safety Assurance	Safety Performance Monitoring & Measurement	PMM01	Quality criteria for safety metrics/indicators have been defined (e.g., validity, sensitivity, practicality)	59
		PMM02	There are requirements to ensure the accuracy and reliability of safety-related data	74
		PMM03	Safety performance metrics/indicators have been defined	100
		PMM04	Periodicities for monitoring all safety performance metrics/indicators are defined	85

		PMM05	Safety performance targets have been defined	89
		PMM06	Limits and thresholds for safety performance indicators have been defined	77
		PMM07	Safety performance is monitored through all predefined metric(s)/indicator(s)	86
		PMM08	All safety metrics/indicators meet all predetermined quality criteria	74
		PMM09	All safety performance data used for monitoring meet all predefined accuracy/reliability requirements	76
		PMM10	All safety performance metrics/indicators are always monitored with the predetermined periodicity	82
		PMM11	Safety performance metrics/indicators cover all safety objectives	78
		PMM12	Defined limits/thresholds of all safety metrics/indicators reflect all safety objectives	62
		PMM13	All safety performance information is included in the first safety communication activities following its compilation	69
		PMM14	All relevant data from mandatory reporting are used in safety performance monitoring	90
		PMM15	All relevant data from voluntary reporting are used in safety performance monitoring	86
		PMM16	There is a method for evaluating the effectiveness of risk controls	70
		PMM17	Review dates are defined for assessing the effectiveness of all risk controls	87
		PMM18	The effectiveness of risk controls is always evaluated according to the predetermined method	86
		PMM19	Review dates for assessing effectiveness of risk controls are always later than the ones defined for the full implementation of risk controls	83
		PMM20	The effectiveness of risk controls is always evaluated before the review date is due	80
	The Management of Change	CM01	There is a change management method	89
		CM02	There are applicability criteria for change management	95
		CM03	Due dates of all changes have been defined	68
		CM04	All changes have been approved by the respective management level	88
		CM05	Change management is always performed according to the predefined method	79
		CM06	Change management is performed always when applicability criteria are met	78
		CM07	Change management criteria and procedures are linked with the Risk assessment criteria and procedures	88
		CM08	All resources required for the implementation of changes are available when the changed are implemented	83
		CM09	All changes are communicated before their implementation begins	79
		CM10	All changes are applied before due dates	72
	Continuous Improvement of SMS	CI1	SMS improvement procedures are described in SMS documentation	77
		CI2	SMS is improved according to the predefined procedures	81
		CI3	All individual SMS changes have been approved by senior management	90
		CI4	The need for SMS improvement has been always examined when safety performance targets were not met	81
		CI5	SMS changes have been always approved only after examining and managing possible conflicts with other business objectives/management systems	74
		CI6	All SMS changes have been included in the first safety communication activities following the changes	75
Safety Promotion	Safety Training & Education	TE1	There is a safety training and education curriculum/program	91
		TE2	The periodicity of safety training and education is defined	85
		TE3	Safety training and education is provided to all employees	90
		TE4	All safety training and education follows at least the predetermined curriculum	98

		TE5	Safety training and education is provided within the predefined periodicity	96
		TE6	Safety training and education curricula/programs are customized to all functions/job roles	90
		TE7	Safety training and education curricula/programs are regularly updated with all relevant internal safety information	91
		TE8	Safety training and education curricula/programs are regularly updated with all relevant external safety information	79
	Safety Communication	COM1	Internal sources for feeding safety communication are defined	94
		COM2	External sources for feeding safety communication are defined	70
		COM3	Periodicity for safety communication is defined	68
		COM4	Safety communication includes all recent and relevant information from predefined internal sources	82
		COM5	Safety communication includes all recent and relevant information from predefined external sources	83
		COM6	Sufficient background that can support the explanation of all external safety information is collected (e.g., in case that a person wants to be informed in more detail).	74
		COM7	Safety communication is provided to all employees	89
		COM8	Safety communication is always provided within predetermined periodicity	89
		COM9	Safety communication is customized to all job functions/roles	59

Appendix E

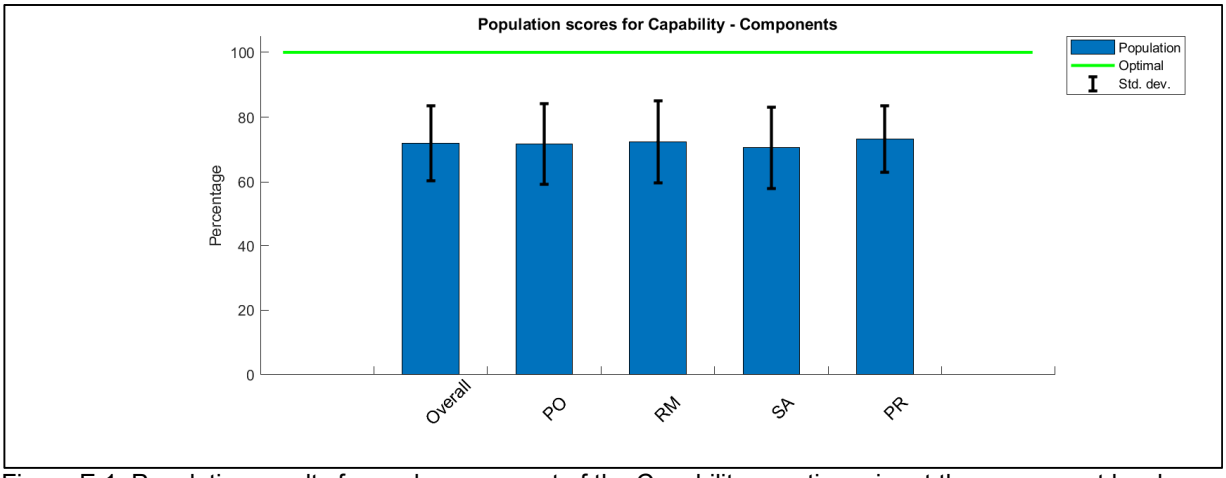

Figure E.1. Population results for each component of the Capability questionnaire at the component level.

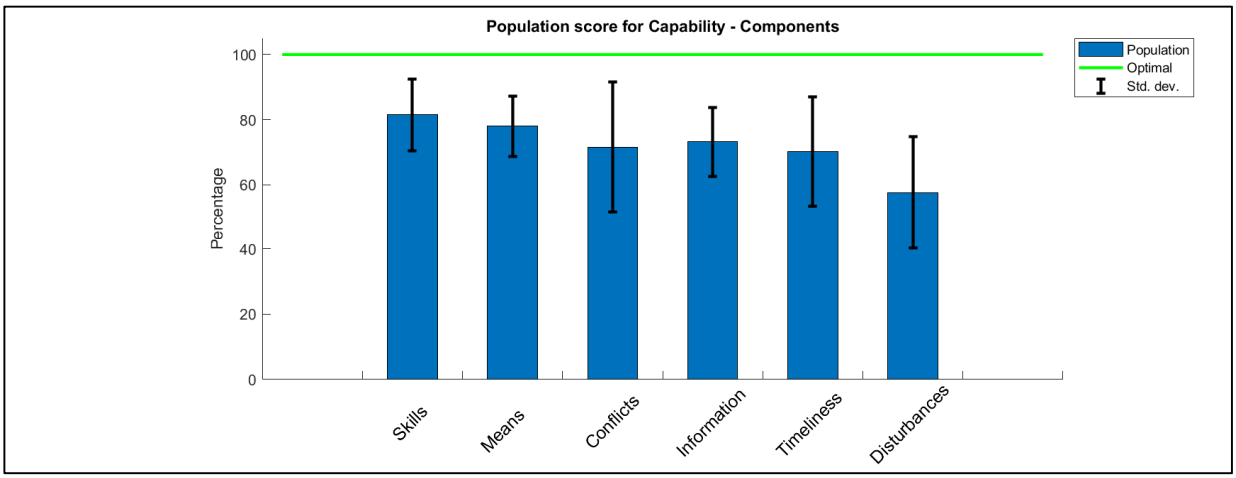

Figure E.2. Population results for each dimension of the Capability questionnaire at the component level.

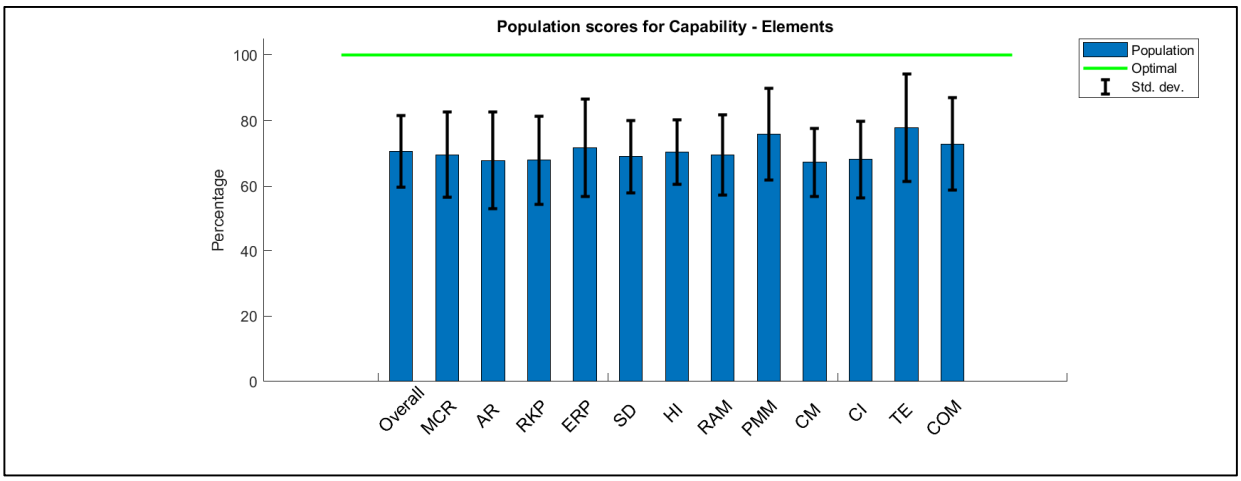

Figure E.3. Population results of the Capability questionnaire at the element level.

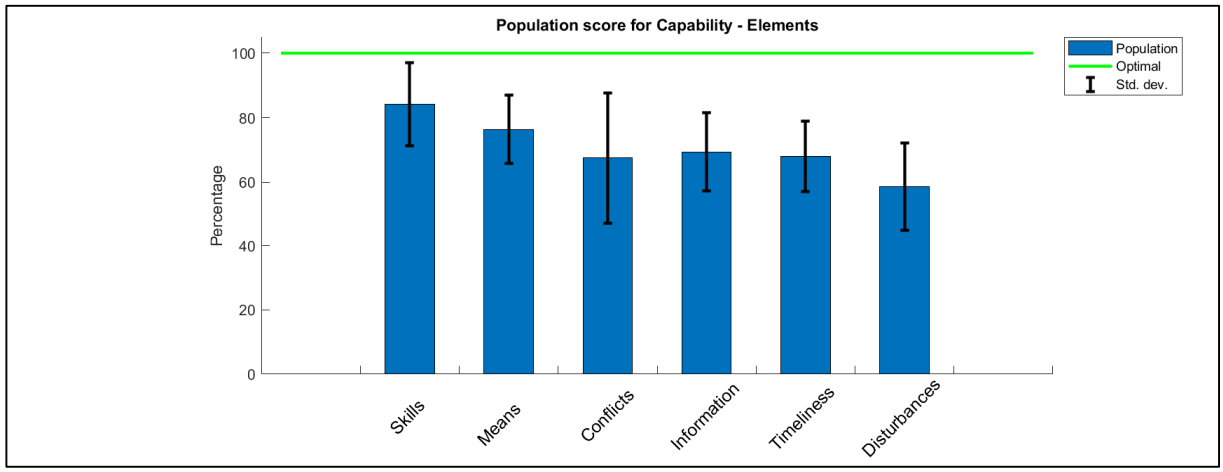

Figure E.4. Population results for each dimension of the Capability questionnaire at the element level.

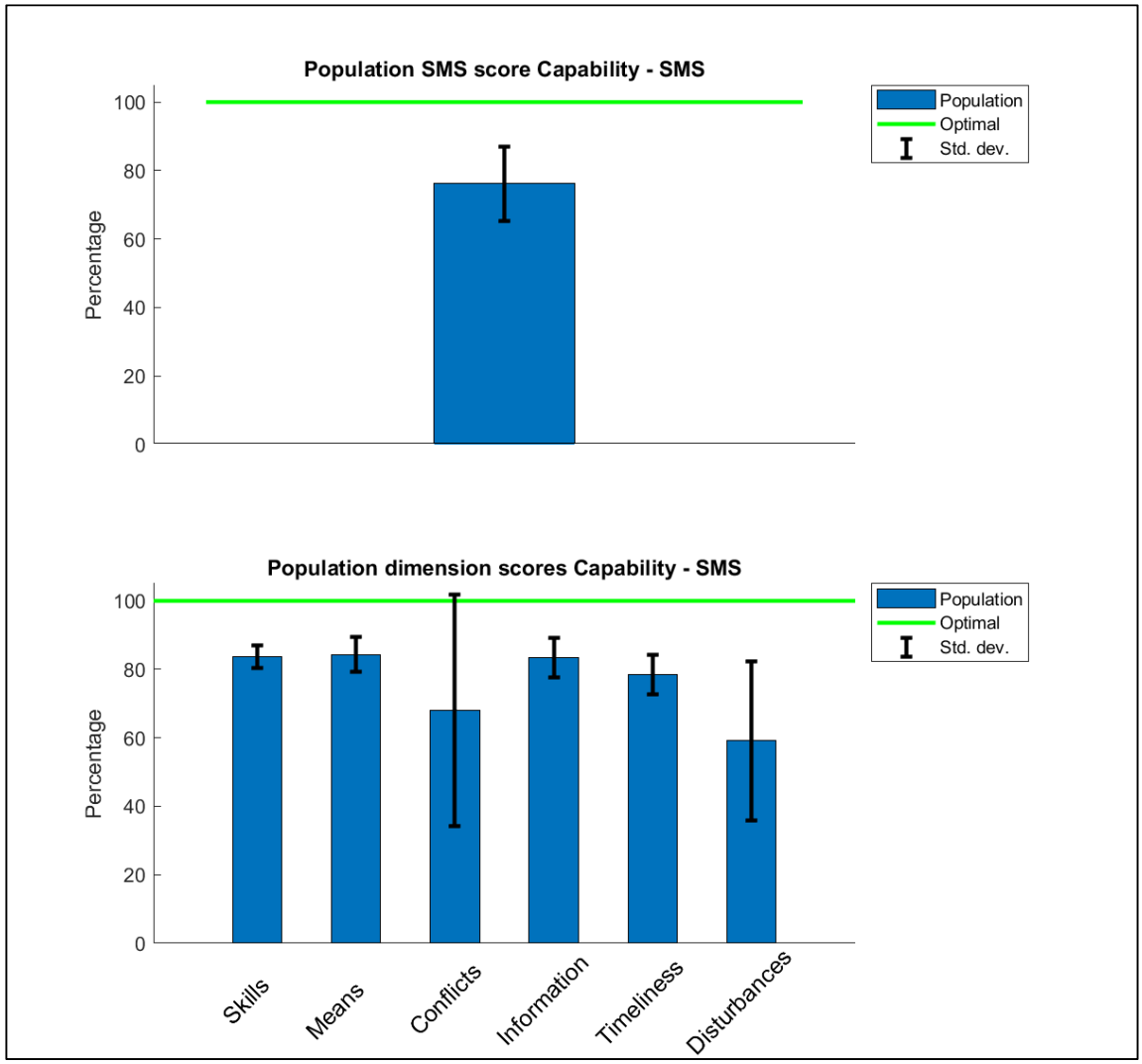

Figure E.52. Population results of the Capability questionnaire at the overall SMS level (top panel) and for each dimension (bottom panel).

Appendix F

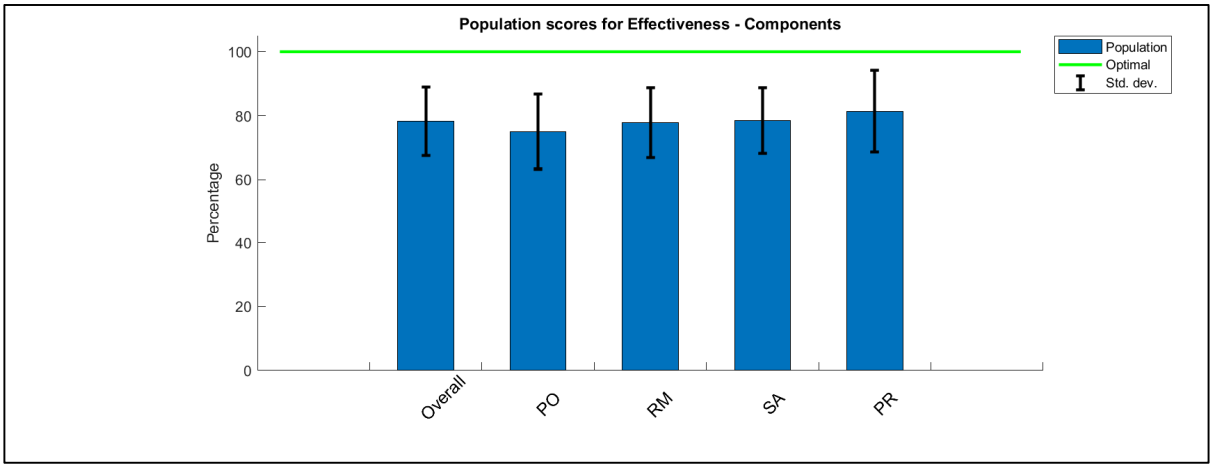

Figure F.1. Population results for each component of the Effectiveness questionnaire at the component level.

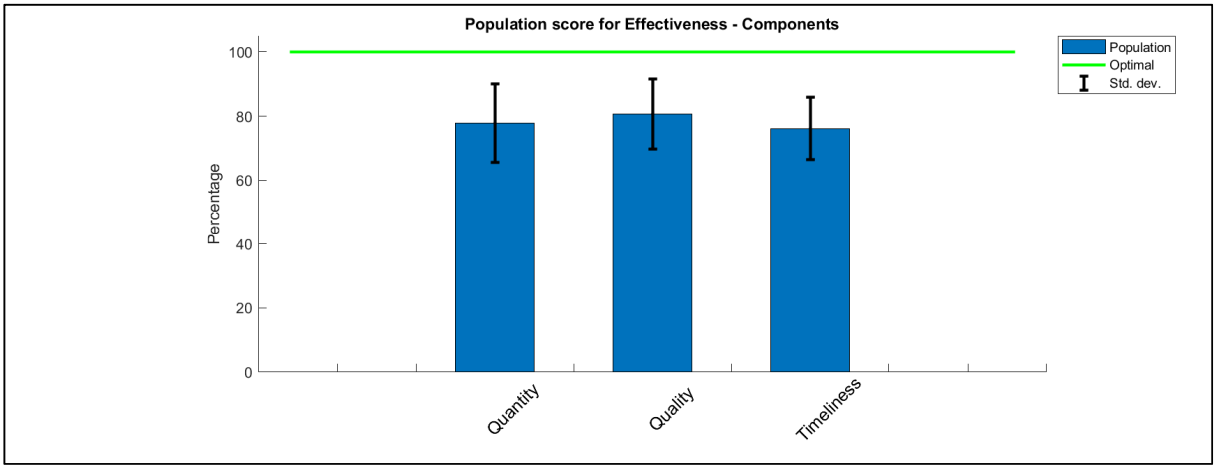

Figure F.23. Population results for each dimension of the Effectiveness questionnaire at the component level.

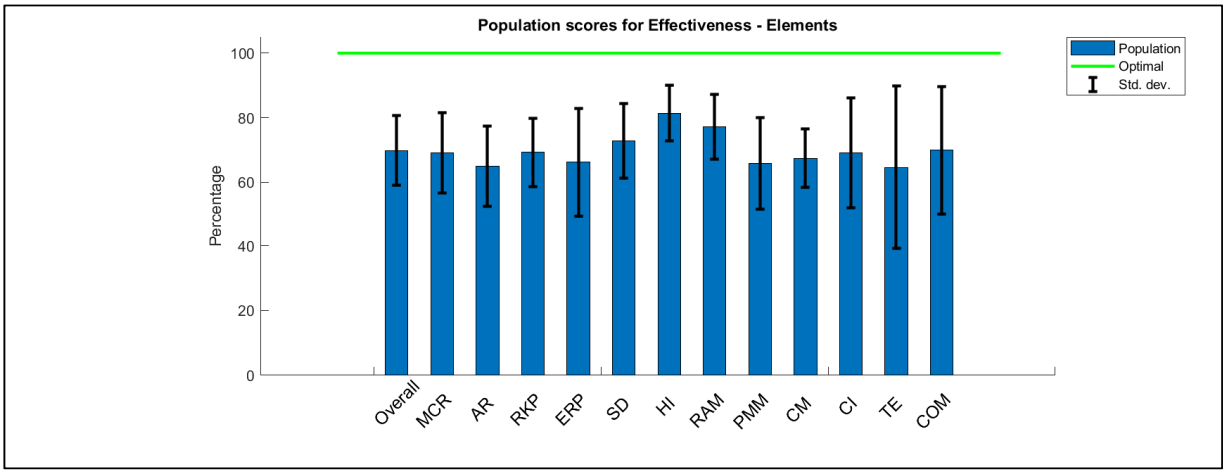

Figure F.3. Population results of the Effectiveness questionnaire at the element level.

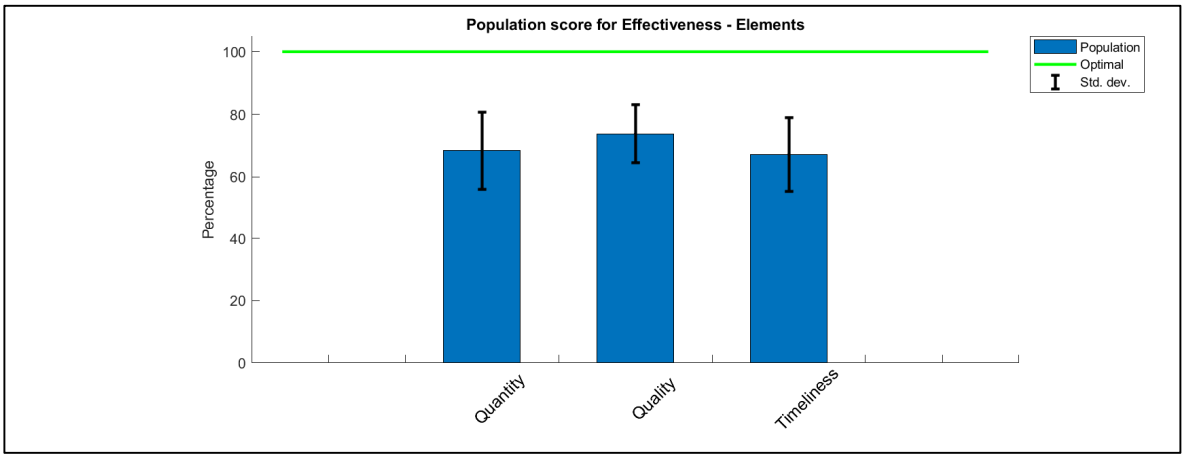

Figure F.4. Population results for each dimension of the Effectiveness questionnaire at the element level.

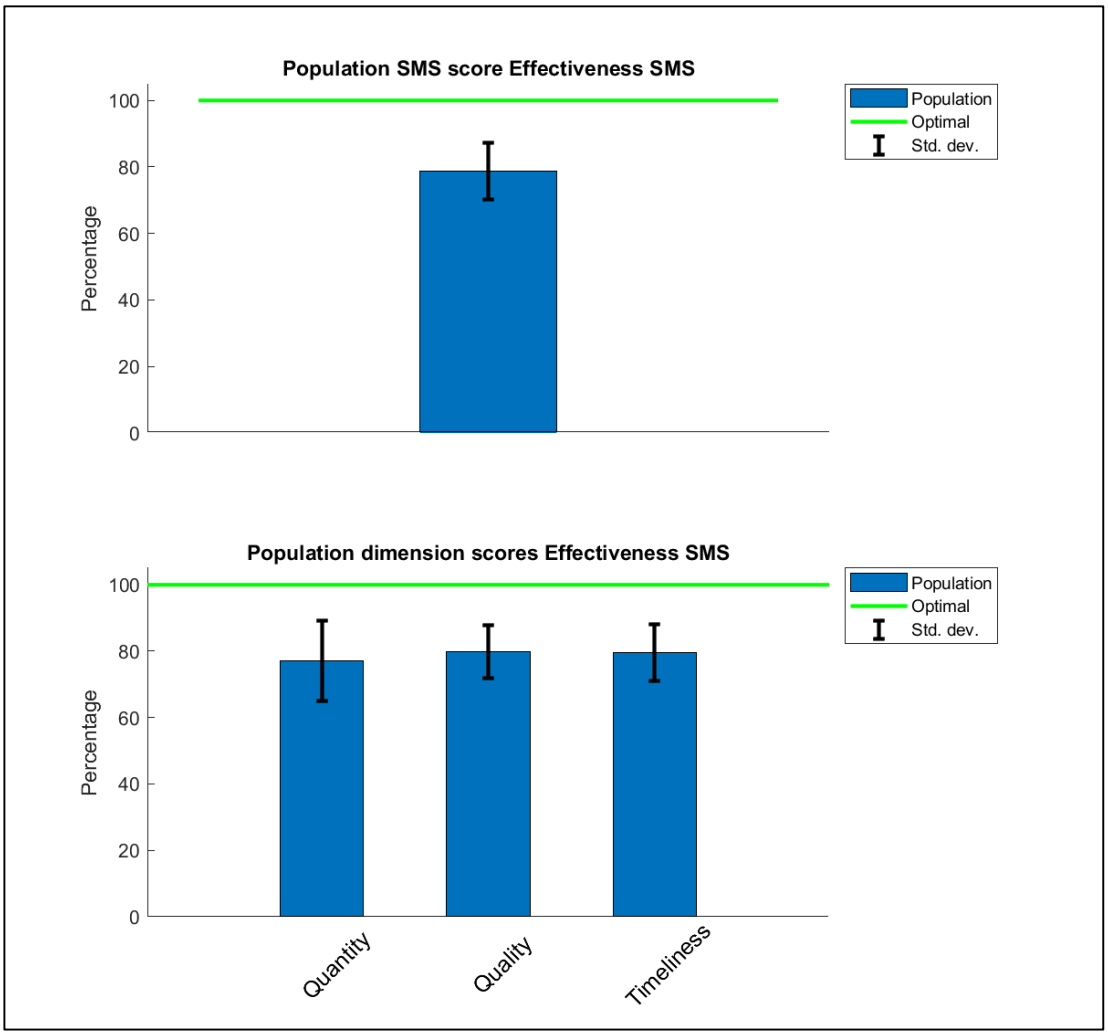

Figure F.5. Population results of the Effectiveness questionnaire at the overall SMS level (top graph) and for each dimension (bottom graph).

Appendix G

Annex G.1: SCP Organizational Plans

Code	Organisational Plans	Yes	Partially	No
G.01-D	There is a written commitment of management towards safety.	100,00%	0,00%	0,00%
G.02-D	Leadership is recognised as important for safety culture development.	96,67%	0,00%	3,33%
G.03-D	Responsibilities & accountabilities for safety have been defined across all management areas.	88,00%	12,00%	0,00%
G.04-D	The safety department is responsible for safety planning ("A basic element of the safety management system (SMS) that enables the setting of organisation's safety objective and targets, as well as the identification of the necessary means and resources for their achievement").	96,00%	0,00%	4,00%
G.05-D	The organisation requires employees' engagement into:			
G.05-D.1	Initial planning activities	69,33%	0,00%	30,67%
G.05-D.2	Monitoring activities	98,67%	0,00%	1,33%
G.05-D.3	Improvement activities	88,67%	0,00%	11,33%
G.06-D	The need for continuous improvement - regardless of past successes - is acknowledged.	100,00%	0,00%	0,00%
G.07-D	A risk management framework exists to be used in decisions about changes and plans.	90,00%	0,00%	10,00%
G.08-D	The risk management framework is tailored to each organizational level, department or work instruction.	39,89%	37,44%	22,67%
G.09-D	The organization acknowledges that buffers are needed to cope with unexpected operational events.	84,00%	0,00%	16,00%
G.10-D	There is a policy for rewarding exceptional contributions to safety (such as new ideas, voluntary participation in safety plans etc., but not daily performance)	30,89%	0,00%	69,11%
G.11-D	The organization has a platform to facilitate internal communication.	97,00%	0,00%	3,00%
G.12-D	The organization facilitates a questioning attitude (e.g.peer reviews, brainstorm sessions, formalised feedback)	82,67%	0,00%	17,33%
G.13-D	The organization provides guidance in effective conflict management.	53,78%	0,00%	46,22%
G.14-D	The organization has communication channels with society.	90,33%	0,00%	9,67%
G.15-D	The organization has communication channels with the authorities.	100,00%	0,00%	0,00%
G.16-D	The organization has communication channels with other sectors.	90,00%	0,00%	10,00%
J.1-D	Documented definitions of "unacceptable" and/or "acceptable" behaviours are defined, accompanied by examples, assumptions, indications, and required evidence etc.	67,33%	0,00%	32,67%
J.2-D	The decision for evaluating unacceptable behaviour is required to be made and agreed by a team including peers.	54,67%	0,00%	45,33%
J.3-D	Rights and duties of employees regarding safety occurrences are described.	93,33%	0,00%	6,67%
J.4-D	The description includes a list of indicative measures and the cases that these might apply to.	56,00%	0,00%	44,00%
J.5-D	Guidelines for prevention of stigmatization of practitioners after an adverse outcome are provided.	50,78%	0,00%	49,22%
J.6-D	In case of legal disputes the organization supports individuals legally, financially, and/or psychologically.	80,56%	0,00%	19,44%
F.1-D	The organization recognizes that there may be a difference between rules and regulations and daily activities.	83,44%	0,00%	16,56%
F.2-D	Employees have the right to self-organize their tasks within specific limits.	60,67%	0,00%	39,33%
F.3-D	Emergency stop procedures are accessible to employees when safety is compromised.	84,67%	0,00%	15,33%
F.4-D	Scheduled emergency response exercises are required.	93,33%	0,00%	6,67%
F.5-D	Unscheduled emergency response exercises are foreseen.	77,00%	0,00%	23,00%
R.1-D	There is a policy for safety reporting.	96,67%	0,00%	3,33%
R.2-D	The implications of safety reporting are described.	96,67%	0,00%	3,33%
R.3-D	When it comes to reporting, definitions of who, what, when, where, how/why are provided.	79,33%	19,33%	1,33%
R.4-D	Reporting is voluntary	93,33%	0,00%	6,67%
R.5-D	Reporting is non-punitive	98,67%	0,00%	1,33%
R.6-D	Reporting is confidential	92,86%	0,00%	7,14%
R.7-D	Reporting is user-friendly	96,67%	0,00%	3,33%
R.8-D	Reporting is easily accessible for the reporter	90,00%	0,00%	10,00%
R.9-D	Reporter is provided with timely feedback	82,67%	0,00%	17,33%
I.1-D	A safety information system is in place.	93,33%	0,00%	6,67%
I.2-D	The safety information system is required to be user-friendly.	92,00%	0,00%	8,00%
I.3-D	The safety information system is required to be freely accessible to all employees.	85,33%	0,00%	14,67%
I.4-D	The safety information system is required to include:			
I.4-D.1	reactive safety information (e.g., accident investigation reports)	93,33%	0,00%	6,67%
I.4-D.2	proactive safety information (e.g., trends derived from voluntary reports, safety inspection/audits)	83,33%	0,00%	16,67%
I.4-D.3	internal safety topics (e.g., improvement plans, newly introduced risk controls, safety management documentation and changes)	92,00%	0,00%	8,00%
I.4-D.4	external safety topics (e.g., safety performance of the sector/other organisations, safety initiatives from authorities)	80,67%	0,00%	19,33%

Code	Organisational Plans	Yes	Partially	No
I.5-D	Time is allocated into employees' working schedules for accessing safety information.	46,89%	0,00%	53,11%
I.6-D	Sharing of safety information across the organization through safety activities (safety meetings, workshops etc.) is required.	100,00%	0,00%	0,00%
L.1-D	The need to learn from safety failures (e.g. Safety investigation reports, voluntary reports, audits) is recognised.	100,00%	0,00%	0,00%
L.2-D	During safety investigations, company policy urges the organization to also examine successes relative to the incident.	47,67%	0,00%	52,33%
L.3-D	Documentation urges management and leaders to promote safety successes across the organisation.	77,67%	0,00%	22,33%
L.4-D	General training of all employees about safety management is required.	95,33%	4,67%	0,00%
L.5-D	Internal comparisons (e.g. across departments, units or individual employees) are required.	56,78%	0,00%	43,22%
L.6-D	External comparisons (e.g. with similar companies and other sectors) are required.	71,44%	0,00%	28,56%
L.7-D	Safety information is used to initiate policy and attitude changes.	95,71%	0,00%	4,29%

Annex G.2: SCP Implementation

	Implementation	Median
G.01-I	My commitment towards safety is clearly visible.	4,75
G.02-I	I take the lead in developing a positive safety culture.	4,5
G.03-I	Shifting between target-oriented (setting and achieving goals) and transformational leadership styles (inspiring employees) is important for a good safety culture.	4,75
G.04-I	I am able to carry out my safety responsibilities.	4
G.05-I	The safety department undertakes its responsibility for safety planning.	4,25
G.06-I	I involve employees into:	
G.06-I.1	Initial planning activities?	4
G.06-I.2	Monitoring activities?	4
G.06-I.3	Improvement activities?	4
G.07-I	I continuously aim to improve safety, regardless of past successes.	5
G.08-I	How often do you base decisions, changes and plans on a risk management framework?	4
G.09-I	Is the risk management framework applicable to the needs of your department?	4
G.10-I	I include operational buffers to cope with the unexpected.	4
G.11-I	I reward exceptional contributions to safety (such as new ideas, voluntary participation in safety plans etc., but not daily performance)?	4
G.12-I	I use available platforms to communicate about safety internally.	4
G.13-I	I facilitate a questioning attitude (e.g. peer reviews, brainstorm sessions, formalised feedback).	4
G.14-I	I am equipped to provide effective conflict management.	4
G.15-I	I use available channels to communicate about safety externally.	4
G.16-I	I use available channels to communicate with the authorities.	4
G.17-I	I use available channels to communicate with other sectors.	4
J.1-I	I use documented definitions of "unacceptable" and/or "acceptable" behaviours to evaluate someone's behaviour.	4
J.2-I	Unacceptable behaviour is evaluated by a team including peers.	4
J.3-I	Rights and duties of employees regarding safety occurrences are known.	4
J.4-I	I use this list of indicative measures and the cases when appropriate.	4
J.5-I	In cases of mistakes/errors (within the area of acceptable behaviour) that led to adverse outcomes, support to employees to enable proper functioning in their job is provided.	4
J.6-I	In cases of mistakes/errors (within the area of acceptable behaviour) that led to adverse outcomes, legal/financial/psychological support to employees is provided.	4
F.1-I	I recognize that there may be a difference between rules and regulations and daily activities.	3,25
F.2-I	To what extent employees have the right to self-organize their tasks?	4
F.3-I	Emergency stop procedures are used by employees when safety is compromised.	4
F.4-I	Scheduled emergency response exercises are executed.	4
F.5-I	Unscheduled emergency response exercises are executed.	3
R.1-I	Safety occurences, regardless their severity, are reported by your employees.	4
R.2-I	The implications of reporting are known to the employees.	4,25
R.3-I	All contextual information is provided in the report.	4
R.4-I	Issues are reported voluntarily.	4
R.5-I	Reporting is not punished.	5
R.6-I	Contents of the reports remain confidential.	4
R.7-I	Reporting is user-friendly in operation.	4
R.8-I	Reporting is easily accessible in operation.	4
R.9-I	Reporter is provided with timely feedback.	4
I.1-I	How frequently do you and your employees use the safety information system?	4
I.2-I	The safety information system is user-friendly.	4
I.3-I	The safety information system is freely accessible to all employees.	4
I.4-I	How frequently do you use the following types of safety information:	
I.4-I.1	reactive safety information (e.g., accident investigation reports)	4
I.4-I.2	proactive safety information (e.g., trends derived from voluntary reports, safety inspection/audits)	4
I.4-I.3	internal safety topics (e.g., improvement plans, newly introduced risk controls, safety management documentation and changes)	4
I.4-I.4	external safety topics (e.g., safety performance of the sector/other organisations, safety initiatives from authorities)	4
I.5-I	Time is allocated into employees' working schedules for accessing safety information.	4
I.6-I	Safety information is shared across your employees through dedicated safety activities (safety meetings, workshops, etc.)	4
L.1-I	Information from safety failures (e.g. safety investigation reports, safety audits, voluntary reports) is used to improve learning.	4
L.2-I	During safety investigations, the organization also examines successes relative to the incident.	4
L.3-I	Safety successes are shared with the employees.	4
L.4-I	General training of all employees about safety management is provided.	4,25
L.5-I	Lessons are learned from internal comparisons (e.g. across departments, units or individual employees).	4
L.6-I	Lessons are learned from external comparisons (e.g. with similar companies and other sectors).	4
L.7-I	Safety information has been used to initiate policy and attitude changes.	4

Annex G.3: Perception

	Perception	Median
G.1-P	The management thinks 'finishing the work' is more important than safety.	4,5
G.2-P	I do not follow regulations and procedures if they delay the operation.	5
G.3-P	If I make safety a priority at work, I often don't have enough time to complete my activities.	4
G.4-P	Unnecessary risks are taken when carrying out the work.	4
J.1-P	People who violate procedures or regulations are punished.	3
J.2-P	My colleague is open about mistakes he/she makes.	4
F.1-P	I am encouraged to draw attention to safety problems.	4,5
I.1-P	Safety cannot be improved further in my field of work.	4
I.2-P	Safety is always discussed during team meetings.	4
L.1-P	After an incident, management takes action to prevent a repeat of the incident.	4