Institutionalisation, Capability and Effectiveness of Aviation Safety Management Systems

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

[1] Aviation Academy, Amsterdam University of Applied Sciences, The Netherlands
[2] Netherlands Aerospace Centre, The Netherlands

ABSTRACT

In the frame of an on-going 4-years research project, the Aviation Academy Safety Management Systems (AVAC-SMS) metric for the self-assessment of aviation Safety Management Systems (SMS) was designed based on the Safety Management Manual of the International Civil Aviation Organization and in cooperation with knowledge experts and aviation companies. The particular metric evaluates three areas, namely (1) the degree of *institutionalisation* of SMS (design and implementation of processes), (2) the extent of managers' *capability* to deliver the SMS processes, and (3) the employees' perceived *effectiveness* of the SMS-related deliverables. The metric concludes with a score per area and per SMS component/element assessed, and it is scalable to the size and complexity of each organisation. Results of a survey at 18 aviation companies did not show statistically significant differences in their SMS scores across all three assessment areas but revealed a distance between the area of Institutionalization and the areas of Capability and Effectiveness. Also, differences were detected regarding the scores per SMS component and element within and across companies and assessment areas. The various assessment options offered for the AVAC-SMS metric accommodates the resources each SME and large company can invest in the application of the metric. Even the lowest level of resolution of the SMS metric can trigger companies to investigate further their weaker areas and foster their SMS-related activities. Therefore, the AVAC-SMS metric is deemed useful to organisations that want to self-assess their SMS and proceed to comparisons amongst various functions and levels and/or over time.

Keywords: Safety Management System; SMS Assessment; Work-as-Imagined; Work-as-Done; Safety Metrics.

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[†]. The project responds to three specific needs of the aviation industry as these were expressed during a roundtable in September 2014 (Aviation Academy, 2014) and confirmed during the first phase of the research (Karanikas, Kaspers, Roelen, Piric, & de Boer, 2016b; Kaspers, Karanikas, Roelen, Piric, & de Boer, in press): (1) Small and Medium Enterprises (SME) lack large amounts of safety-related data to measure and demonstrate their safety performance proactively, (2) large companies might obtain abundant data, but they need safety metrics which are more leading than the current ones and of better quality and (3) 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

---

[*] Corresponding author: +31(0)621156287, n.karanikas@hva.nl, nektkar@gmail.com
[†] http://www.regieorgaan-sia.nl/

without the benefit of large amounts of data and with an emphasis on performance rather than mere compliance (Aviation Academy, 2014). Following the mapping of the current situation through literature review (Karanikas et al, 2016b; Kaspers et al., in press), a survey regarding the currently used metrics (Karanikas et al, 2016a; Kaspers et al., 2016; Kaspers et al., 2017) and the design of new safety metrics (Karanikas et al, 2017), the safety current paper focuses on the metric designed for the self-assessment of Safety Management Systems (SMS) (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 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 organisational levels and functions and indicate areas where the gaps between Work-as-Imagined (a.k.a. WaI) and Work-as-Done (a.k.a. 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 (Leveson, 2011) 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; a more elaborate description of the AVAC-SMS is presented in the next paper section:

- 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

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 organisation/service provider with an established ICAO-based SMS.

2. METHODOLOGY

To test the AVAC-SMS metric through the collection of real-world data and examine its value to detect gaps between the Institutionalisation, Capability and Effectiveness areas of SMS as well differences amongst companies, we transferred the respective questionnaires to the Qualtrics platform[‡] and invited aviation companies to participate in the surveys. In total, 18 large and SME companies participated at least in one of the questionnaires per SMS assessment area: 14 from Europe, 2 from North America, 1 from Africa and 1 from the Pacific Region. The company types, sizes and numbers participated in the surveys are presented in Table 1. The questionnaires were offered at three different resolution levels yielding a total of nine questionnaires with respective estimated completion times (Figure 1); 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).

Table 1 Sample distribution

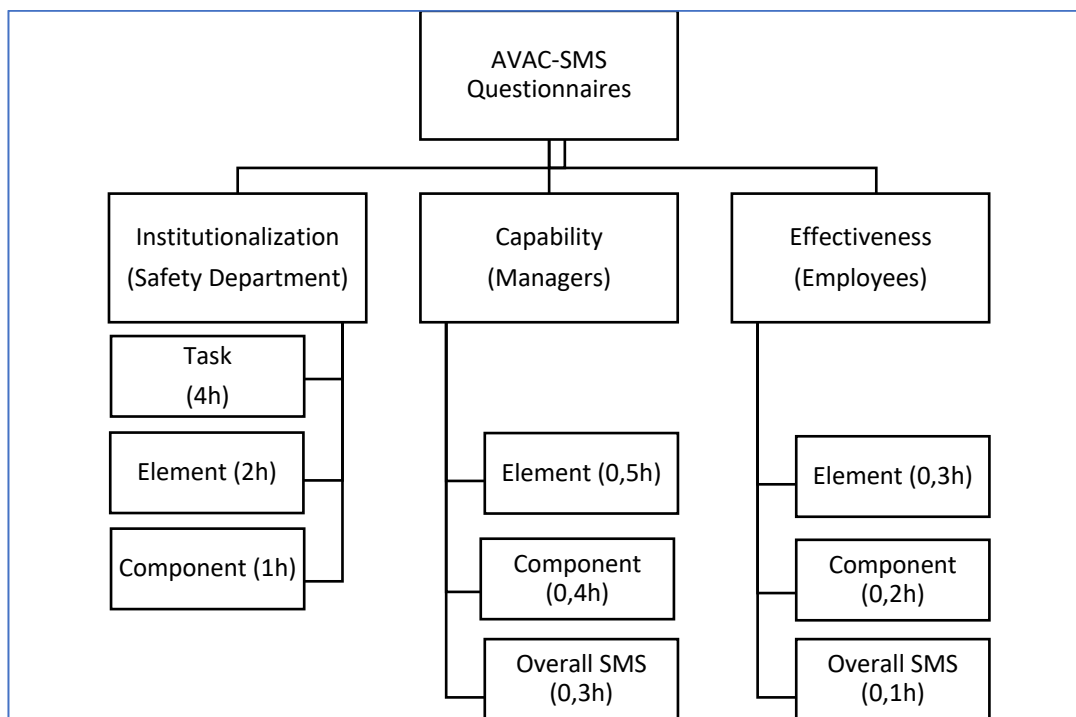| Company type	Size	AVAC-SMS
Air Operator	Large	3
	SME	7
Air Navigation Service Provider	Large	1
Other (airports, ground handlers etc.)	Large	6
	SME	1
Total	Large	10
	SME	8

[‡] www.qualtrics.com

Figure 1: Overview of the AVAC-SMS questionnaires; completion time is reported in brackets

Table 2 presents the distribution of questions for the task level of the institutionalisation dimension. The task level included compliance and implementation questions as well as time and process dependencies. The 149 questions were divided into three aspects: Design (i.e. compliance), Implementation (i.e. realisation 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.

Table 2 Overview of question numbers per component and elements

SMS Institutionalization – Task Level			
	Components	Elements	# Questions/Tasks
SMS	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 Institutionalization – Task Level		
SMS	Total	149

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. realisation 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 3 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.

Table 3 Participation in each of the nine SMS questionnaires

Company Code	Institutionalisation 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)	

Company Code	Institutionalisation 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
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)	

The institutionalisation 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 3, irrespective of the instructions provided, a few companies opted to fill in capability and effectiveness questionnaires at more than one levels, as with the institutionalisation. 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 referred 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 entry were only included if at least 75% of the questions were answered. The calculations were performed as follows (see Appendix A 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 institutionalisation 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 of the respective constructs. 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.

3. RESULTS

3.1. 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 4. As such, only the overall SMS score per questionnaire was used for further calculations. The different scores per SMS area are presented in Table 5; 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.

Table 4 Cronbach's Alpha values for scores aggregated at different SMS levels

	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 5 SMS-level scores per company and questionnaire. Bold entries denote the score that was used as an overall score and corresponds to the result at the highest level of detail

Company Code	Institutionalisation			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**	

Company Code	Institutionalisation			Capability			Effectiveness		
	Task	Element	Component	Element	Component	SMS	Element	Component	SMS
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		

3.2. Institutionalization

The results included in this section are presented graphically in Appendix B. At the component level of assessment (Figure B.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 B.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 B.4), and it provided a similar score for the overall SMS (82,7%). The picture per element (Figure B.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 B.6). The elements which scored higher than or equal to the overall score (Figure B.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 B.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.

3.3. Capability

The results included in this section are presented graphically in Appendix C. The overall SMS capability at the component level of resolution was 72,0% without major differences amongst the scores per component (Figure C.1). The dimensions concerned (Figure C.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 C.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 C.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 C.5), there were insufficient 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.

3.4. Effectiveness

The results included in this section are presented graphically in Appendix D. At the component assessment level (Figure D.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 D.2). The element level concerned (Figure D.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 D.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 D.5).

3.5. 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.

4. DISCUSSION

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 noted that the scores between these areas must be read as follows regarding the Wal-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 organisation.

- 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 value 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 not as an 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 institutionalisation 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 educational 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 area of SMS effectiveness, 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.

5. CONCLUSIONS

The application of the AVAC-SMS metric showed that it has adequate sensitivity to capture any gaps between WaI and WaD amongst different SMS components and levels and across organisations regarding the areas of SMS Institutionalisation, Capability and Effectiveness. Also, the application of the metric revealed interesting differences amongst the various aspects measured: Design, Implementation, Timeliness and Dependencies. 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).

We acknowledge that 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. Also, the metric is designed to collect self-reported data regarding the various dimensions of the SMS areas assessed and, consequently, the findings could have been affected by socially desirable answers. However, we expect that such effects were minimised because the AVAC-SMS metric was communicated to the companies as a self-assessment and not an auditing tool, the participation in the research was voluntary, and we guaranteed the anonymity of the companies as well as their managers and employees. Hence, we believe that the results presented above in combination with the ones communicated to the companies (i.e. company scores and benchmarking against the rest of the sample) can trigger them to investigate further their weaker areas and foster their activities related to SMS. Although based principally on perceptions, the AVAC-SMS metric is deemed as useful to organisations that want to

self-assess their SMS 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 and 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 metric.

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 AVAC-SMS, the data collected was insufficient to perform statistical calculations. Therefore, at this stage, we could not determine whether the AVAC-SMS has any predictive validity. The researchers plan to run a second round of surveys to apply the metric and collect safety/activity data from more organizations. Hence we anticipate that we will be able to test the metric against safety performance and activity figures. Nonetheless, irrespective of the possible associations of the AVAC-SMS metric with safety outcomes, its application and findings communicated in this report are supportive of its usefulness, practicality and potential value for the companies that are interested in assessing their SMS, reveal gaps amongst the specific assessment areas of the metric and get insights into their strong and weak points to improve further the way they manage safety.

Although the AVAC-SMS metric was designed based on STPA, which is a recognised analysis technique, and incorporates the comments of knowledge experts and practitioners stated during the peer-reviews, further research is proposed to perform a more systematic comparison between the particular metric and other SMS assessment tools used in the aviation industry. Such research could collect the views of the industry regarding the relative strengths and weaknesses of the AVAC-SMS metric against others.

ACKNOWLEDGEMENTS

REFERENCES

Aviation Academy. (2014). *Project Plan RAAK PRO: Measuring safety in aviation - developing metrics for Safety Management Systems.* The Netherlands: Amsterdam University of Applied Sciences, Aviation Academy.

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

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

ICAO. (2013). *Doc 9859 Safety Management Manual (SMM)* (3rd ed.). Montreal: International Civil Aviation Organisation.

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

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

Karanikas, N., Selma, P., de Boer, R., Roelen, A., Kaspers, S., & 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, 1(1)*, (pp. 44-57). Amsterdam University of Applied Sciences, AUP Advances. doi:10.5117/ADV2018.1.003.KARA

Kaspers, S., Karanikas, N., Roelen, A., Piric, S., & de Boer, R. (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. (2016). Exploring the Diversity in Safety Measurement Practices: Empirical Results from Aviation. *Proceedings of the 1st International Cross-Industry Safety Conference, Journal of Safety Studies, 2(2)*, (pp. 18-29). Amsterdam, 3 - 4 November 2016. doi:10.5296/jss.v2i2.10437

kaspers, S., Karanikas, N., Roelen, A., Piric, S., van Aalst, R., & de Boer, R. (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).

Levenson, N. (2011). *Engineering a Safer World Systems Thinking Applied to Safety.* Cambridge: MIT Press.

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

APPENDIX A

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 th,en this will be omitted.

Institutionalisation – Task level

Intermediate calculations

Task

Element

Component

Overall SMS

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

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

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

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

SMS Scores

Task

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

Element

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

Component

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

Overall SMS

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

Institutionalisation, 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

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

Component

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

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

$$= \sqrt{\sum_{i=1}^{n_e} (ne_{c_i} \cdot n_q \cdot 100)^2}$$

Overall SMS

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

Institutionalisation, Capability, and Effectiveness – Component level

Intermediate calculations

Component

Overall SMS

$$s_{c_i} = \sum_{q_j \in c_i} resp_{q_j}$$

$$m_{c_i} = \sum_{e_j \in c_i} m_{e_j}$$

$$r_{c_i} = \frac{1}{ne_{c_i}} \sum_{e_j \in c_i} r_{e_j}$$

$$d_{c_i} = m_{c_i} - s_{c_i}$$

$$s_{sms} = \sum s_{c_i}$$

$$m_{sms} = \sum c_i$$

$$r_{sms} = \frac{1}{n_c} \sum r_{c_i}$$

$$d_{sms} = m_{sms} - s_{sms}$$

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

$$D_{sms} = 1 - r_{sms}, M_{sms} = 1$$
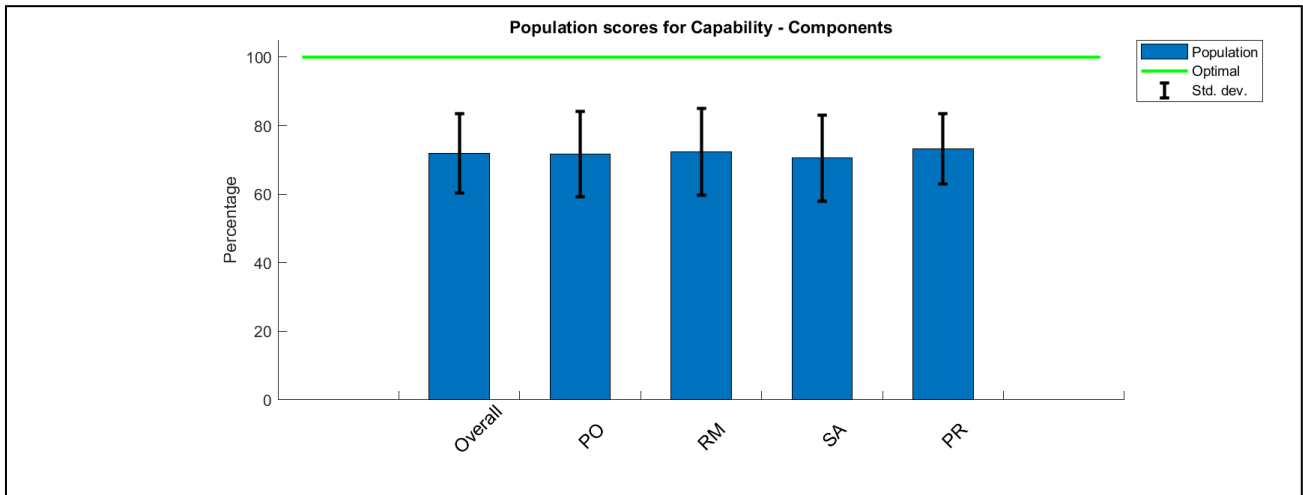
APPENDIX B
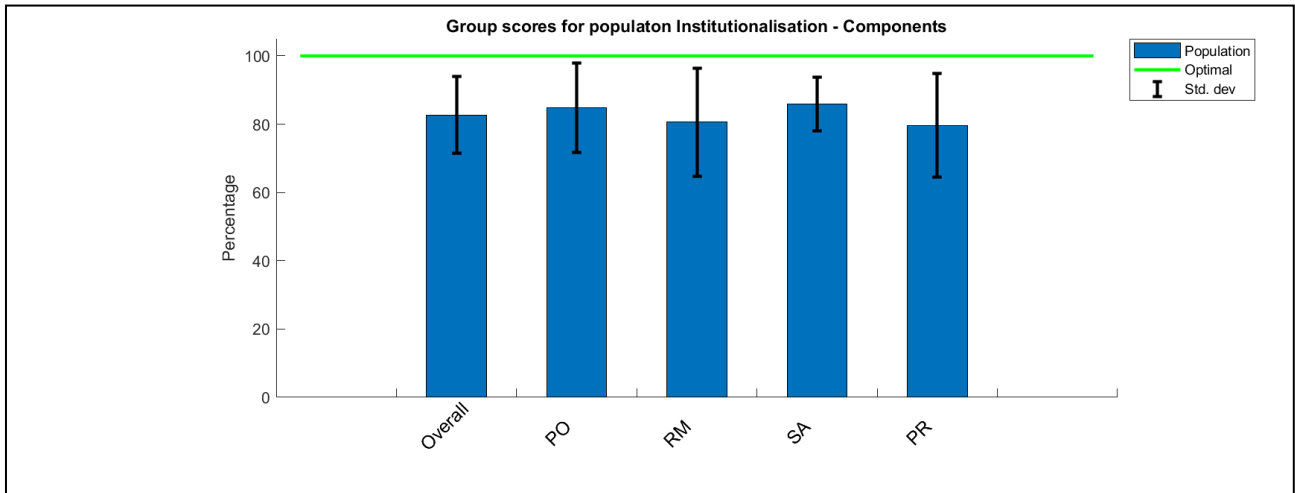
Figure B.1: Population results for each component of the Institutionalisation questionnaire at the component level
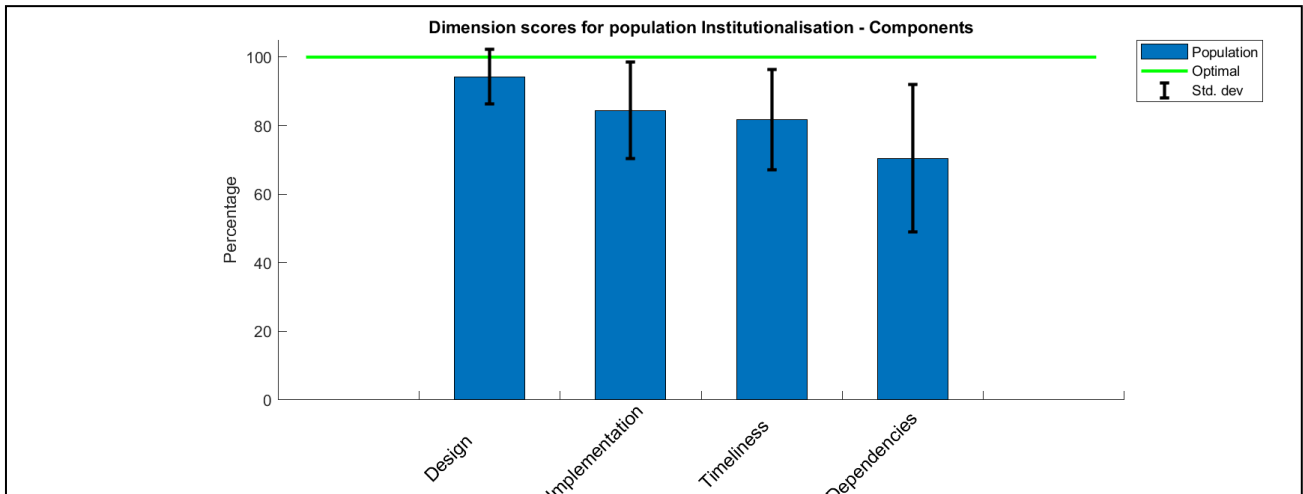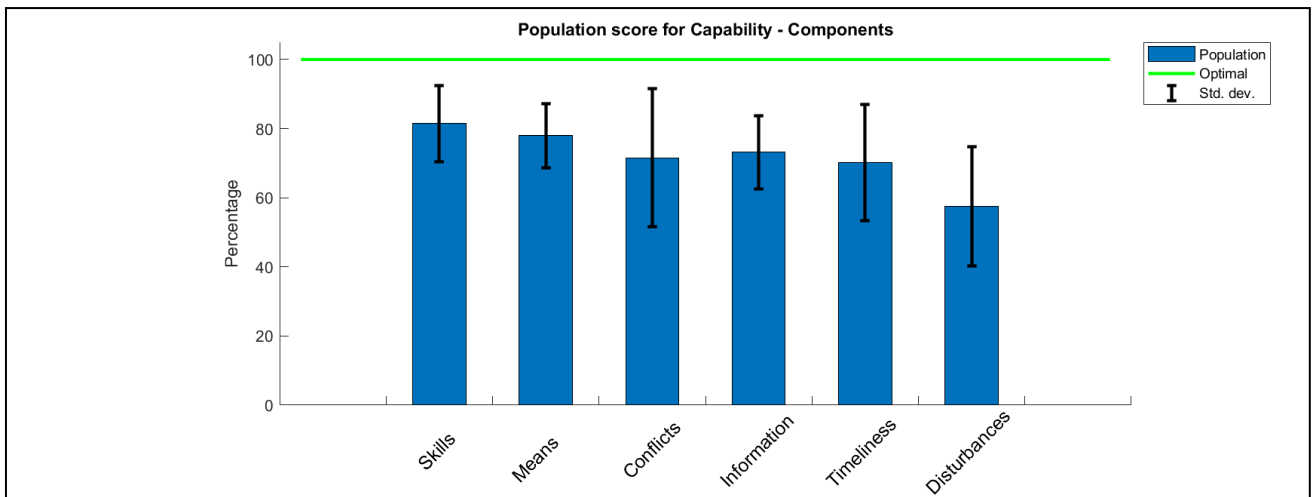

Figure B.2: Population results for each dimension of the Institutionalisation questionnaire at the component level
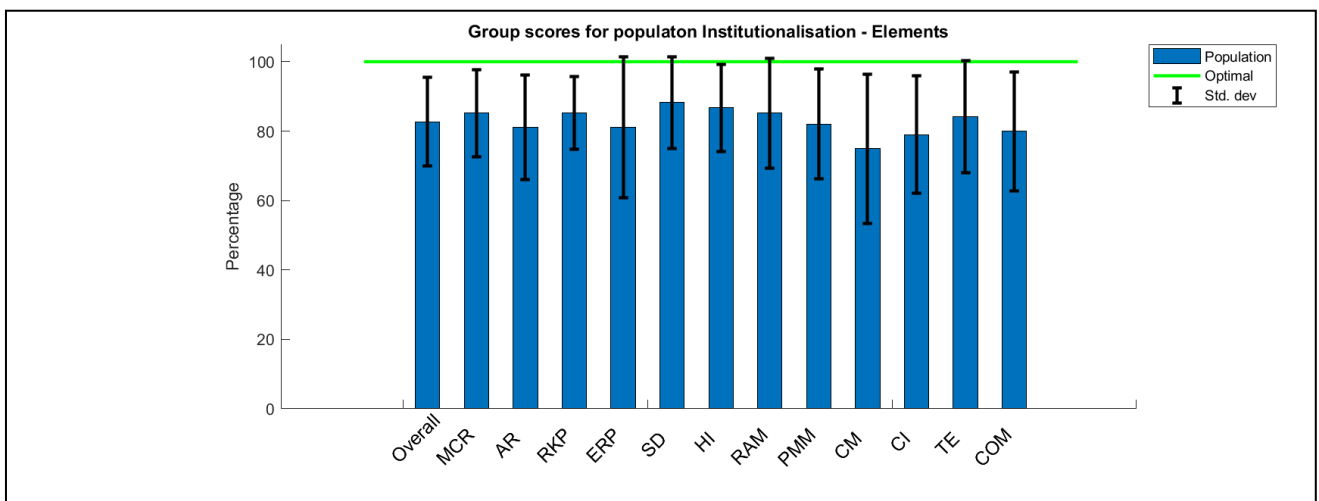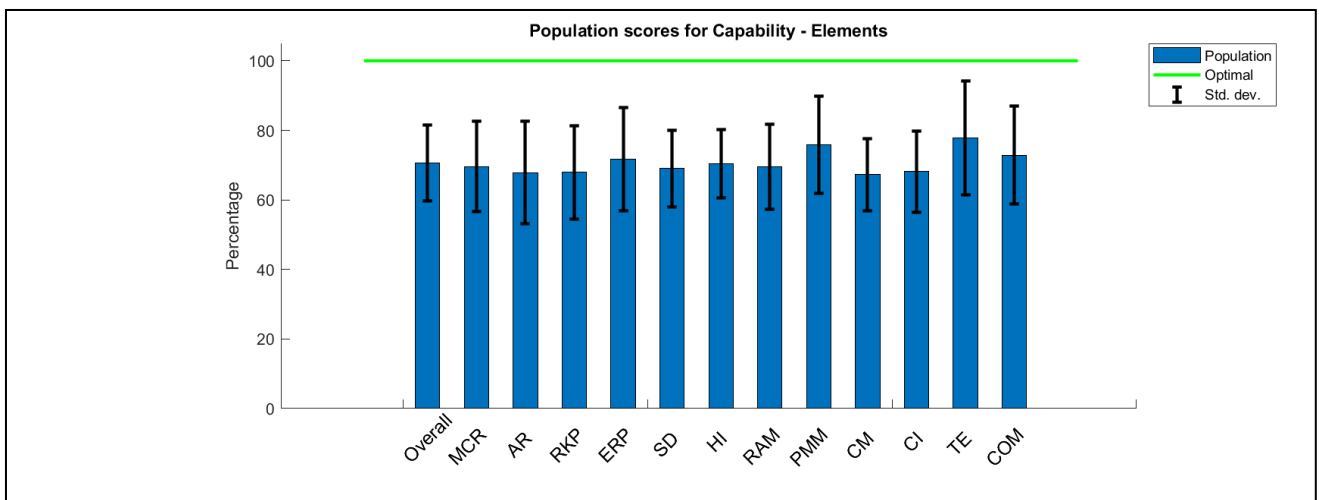

Figure B.3: Population results for each element of the Institutionalisation questionnaire at the element level
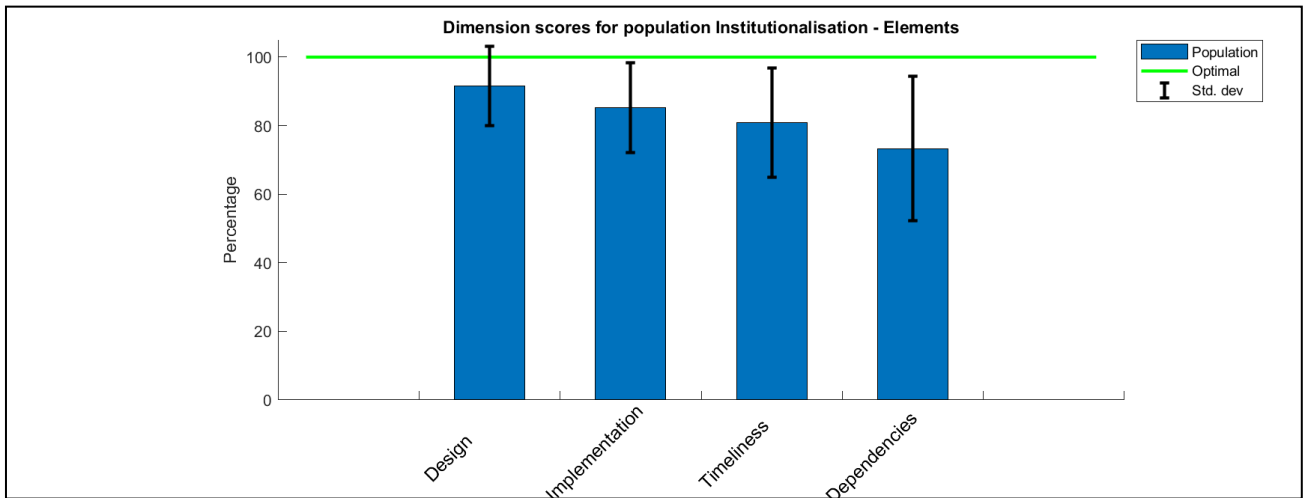
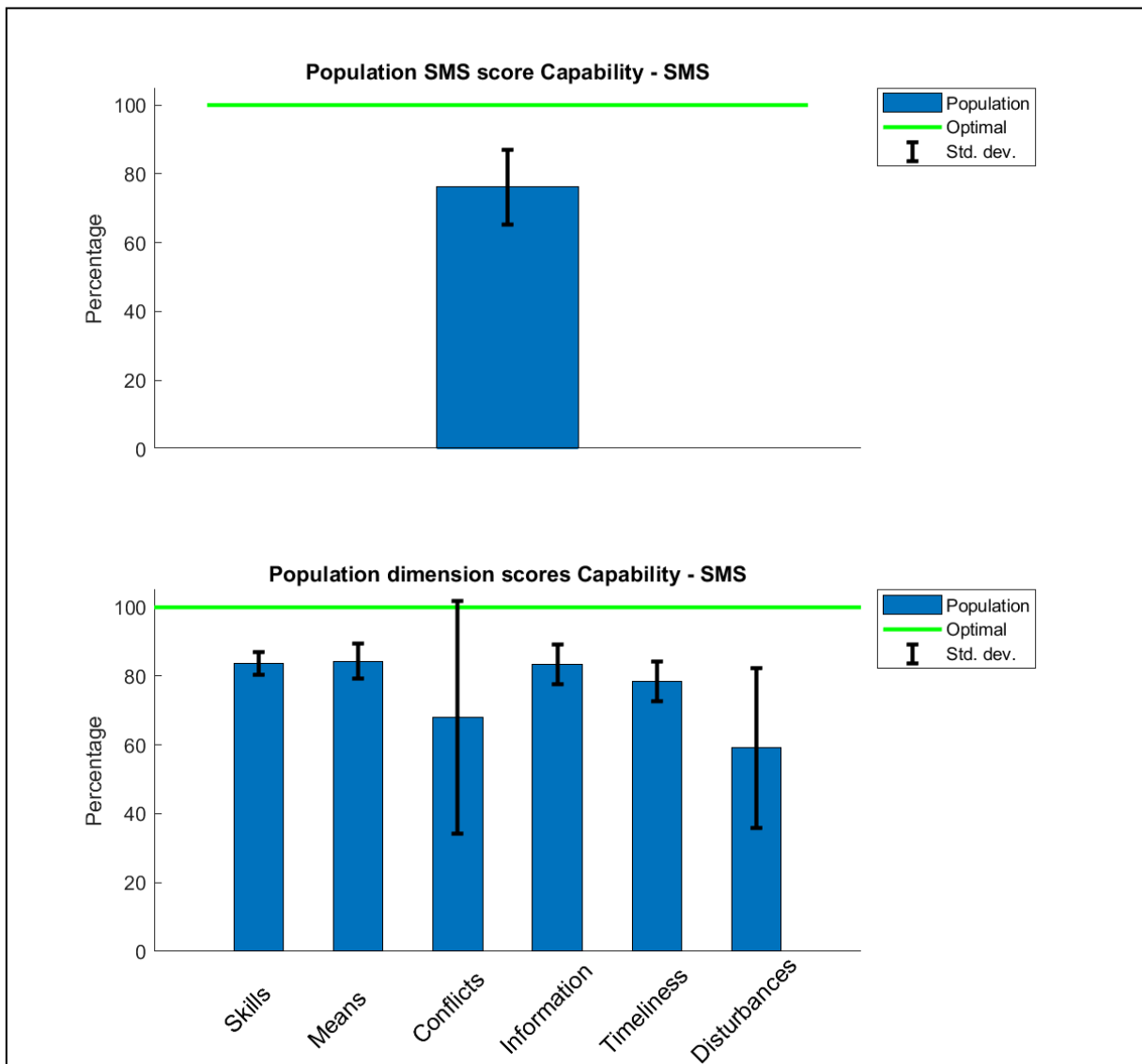Figure B.4: Population results for each dimension of the Institutionalisation questionnaire at the element level
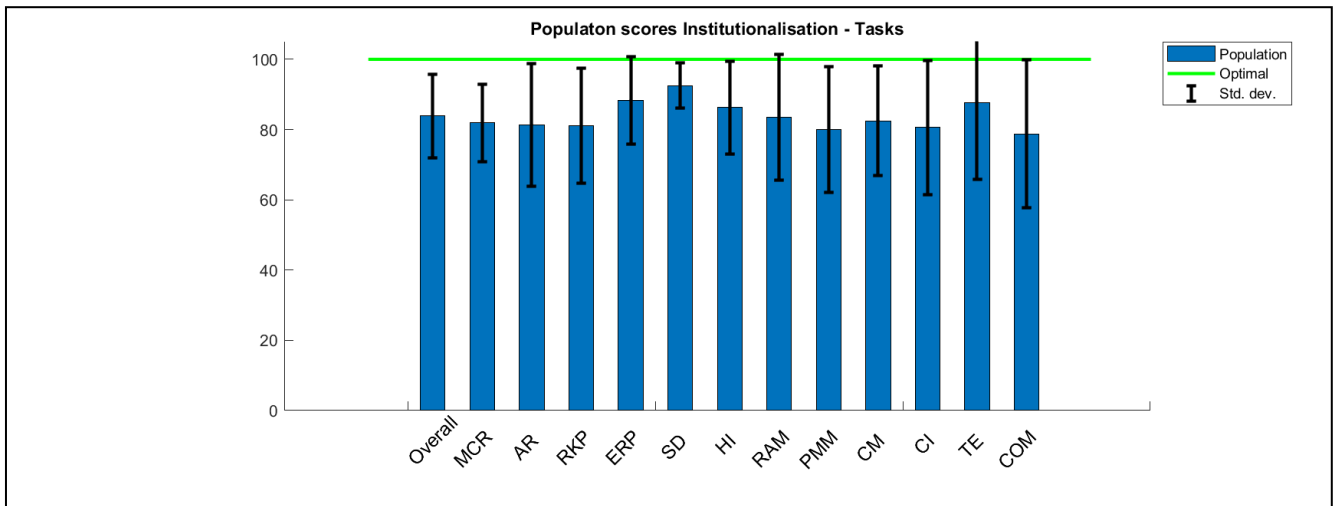
Figure B.5: Population results of the Institutionalisation questionnaire at the level of tasks grouped by element
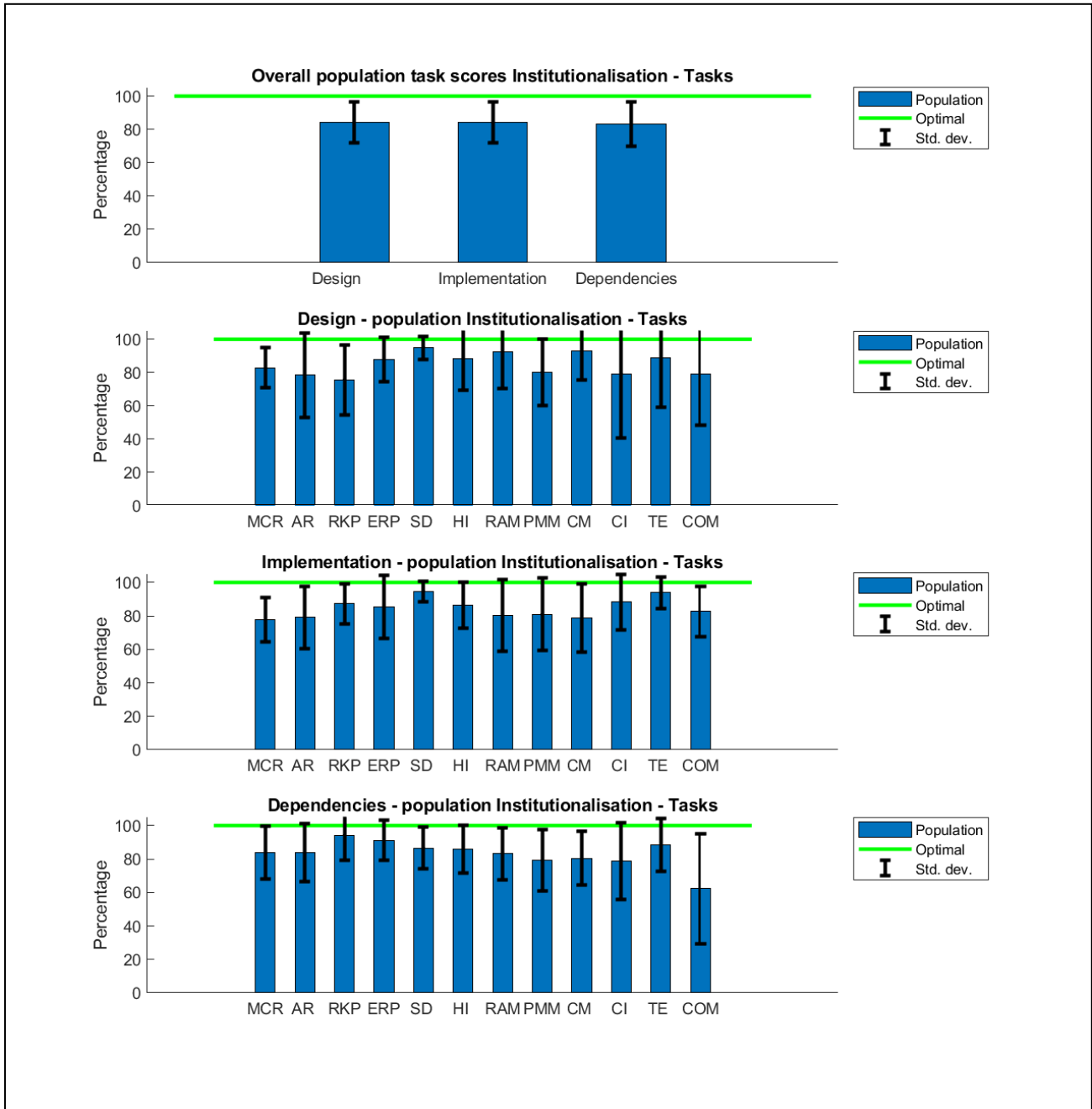
Figure B.6: Population results for the Institutionalisation questionnaire at the level of tasks grouped by aspect (top panel) and each aspect across elements (bottom three panels)
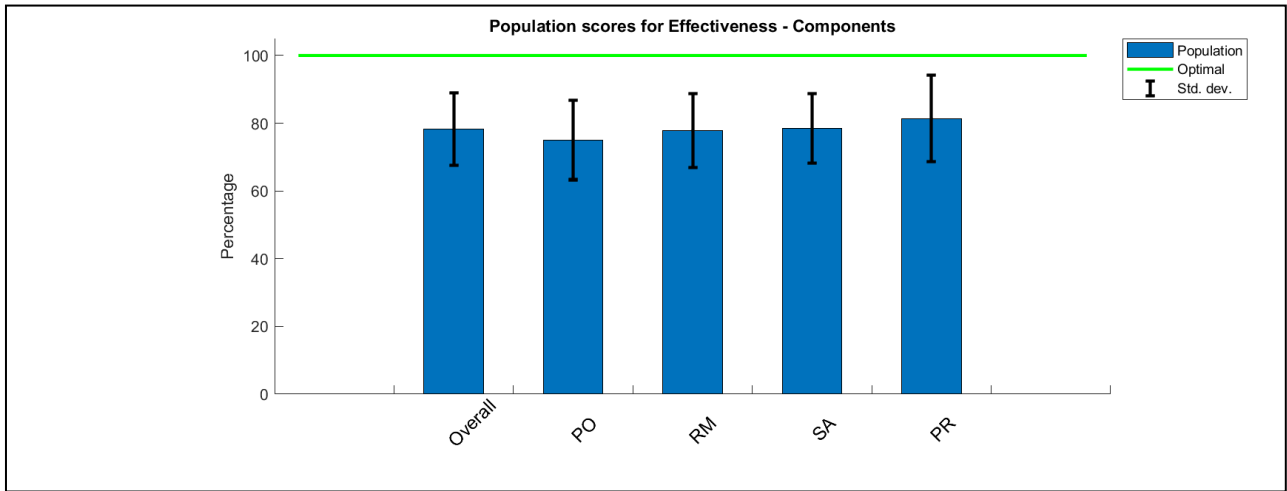
APPENDIX C

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

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

Figure C.3: Population results of the Capability questionnaire at the element level
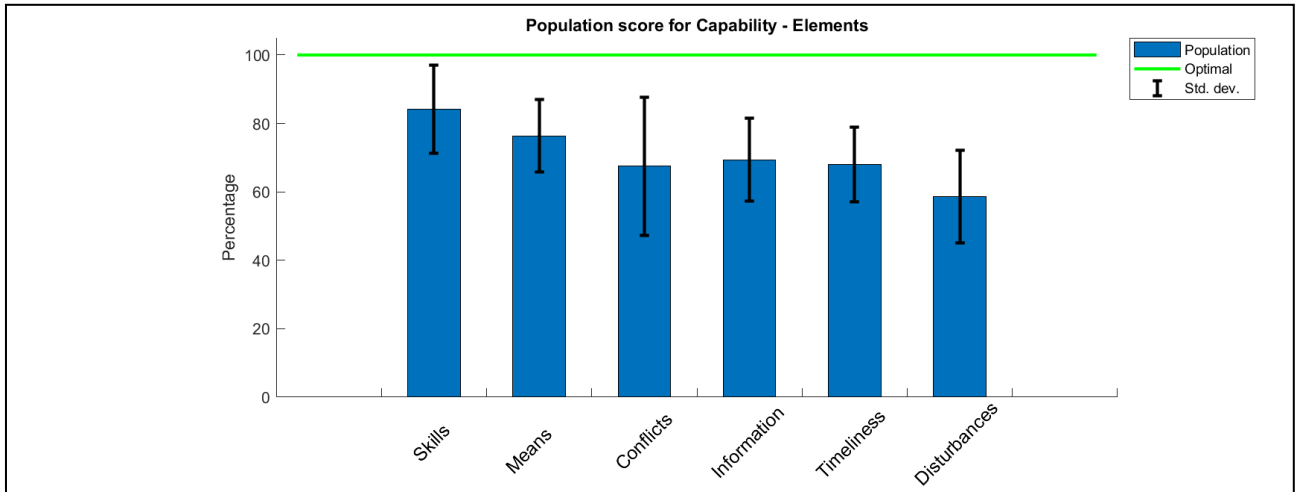
Figure C.4: Population results for each dimension of the Capability questionnaire at the element level

Figure C.51: Population results of the Capability questionnaire at the overall SMS level (top panel) and for each dimension (bottom panel)

APPENDIX D

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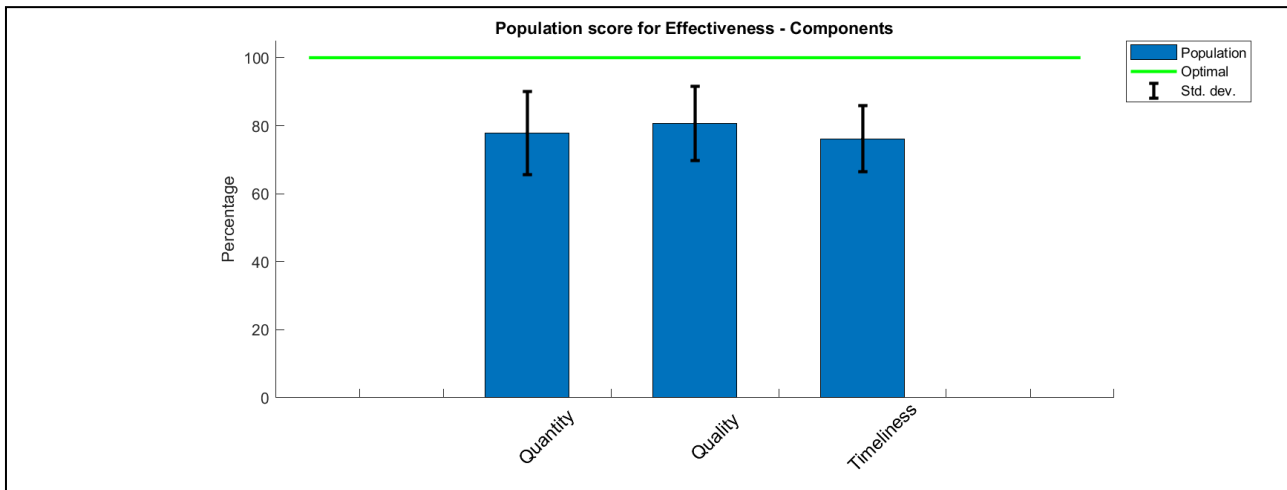

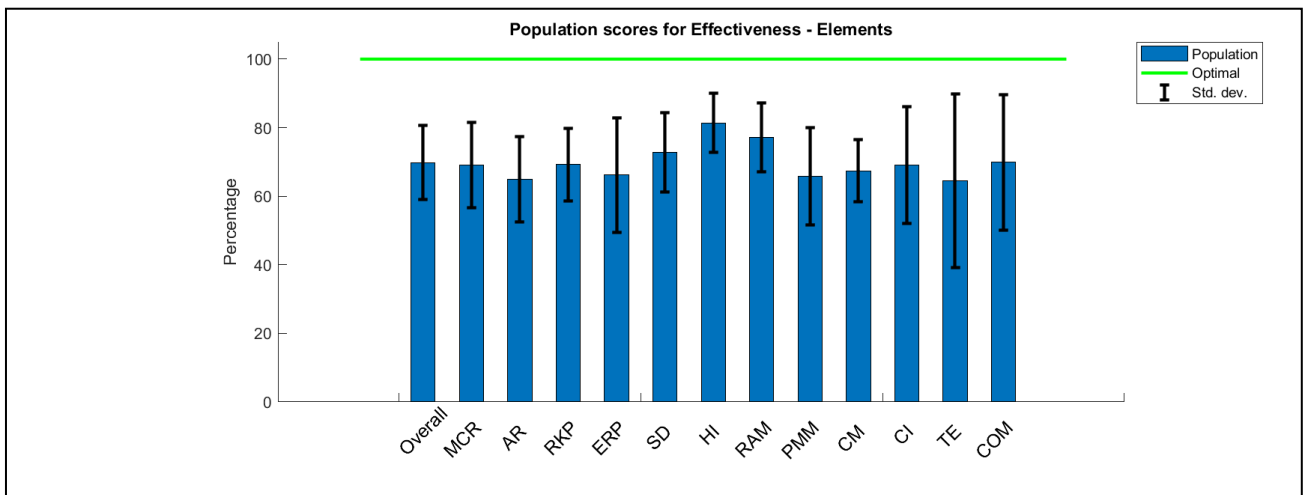Figure D.22: Population results for each dimension of the Effectiveness questionnaire at the component level

Figure D.3: Population results of the Effectiveness questionnaire at the element level
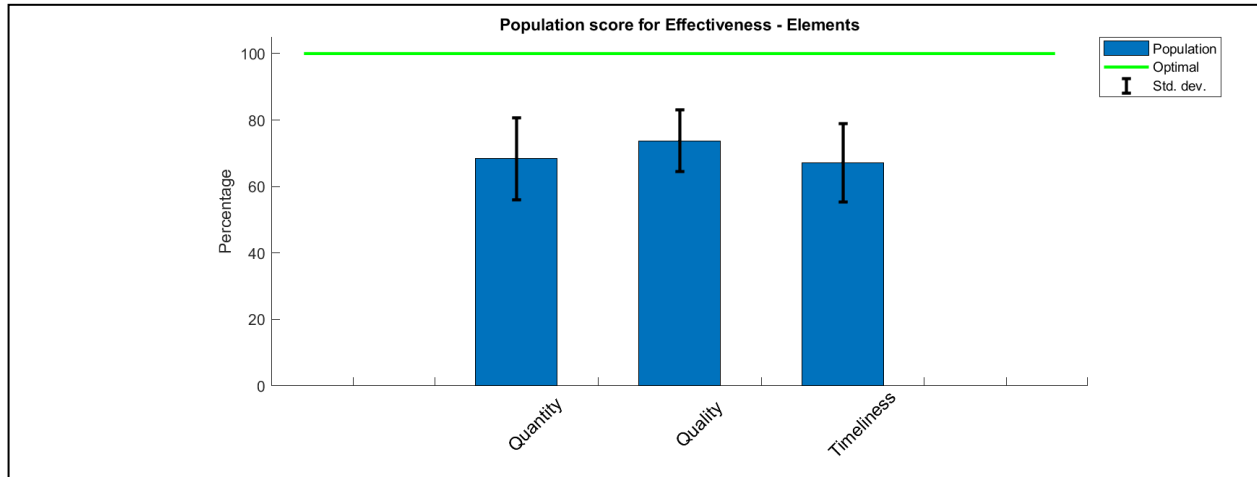
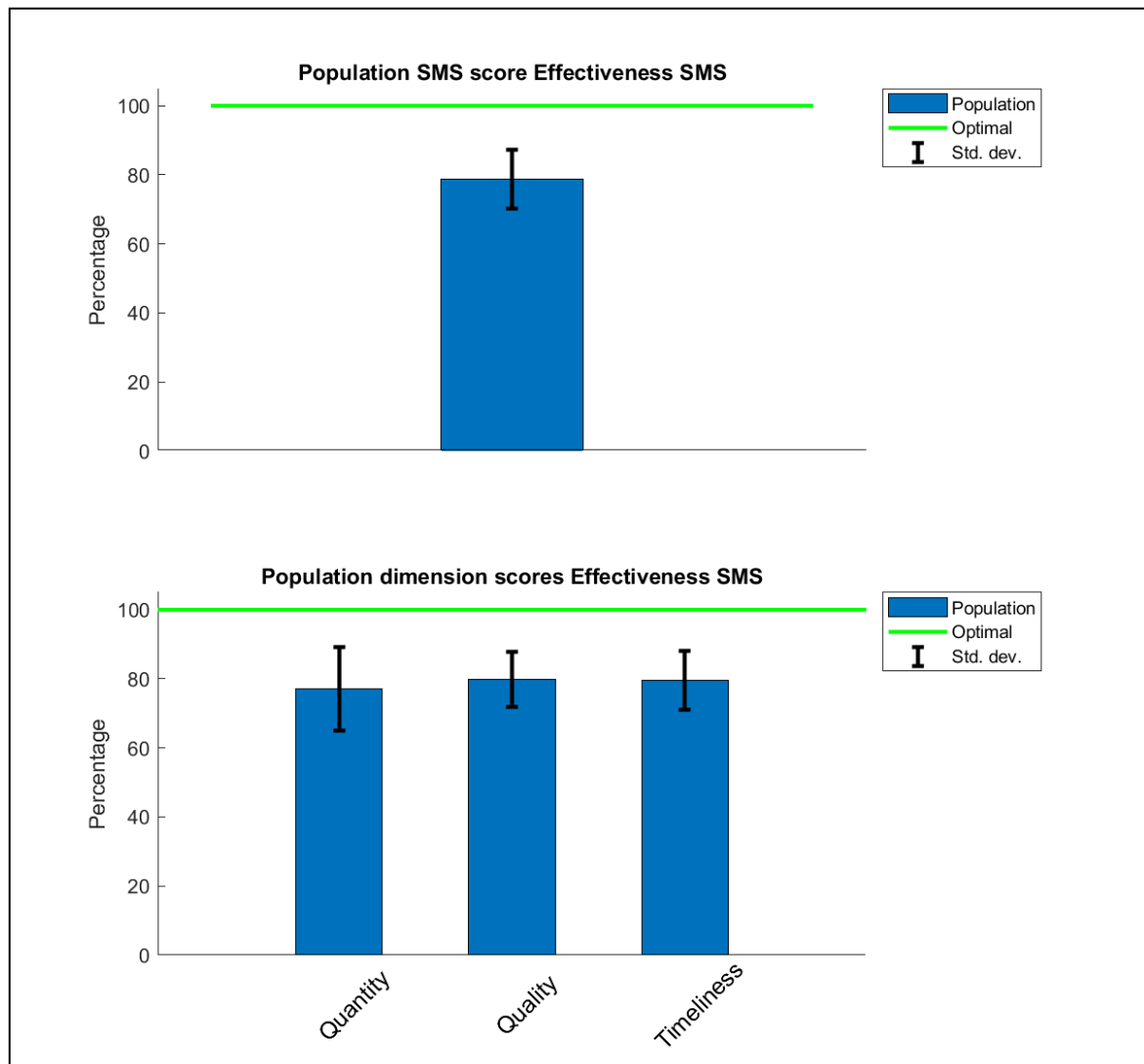Figure D.4: Population results for each dimension of the Effectiveness questionnaire at the element level

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