DEVELOPING A SAFETY CULTURE MEASUREMENT TOOLKIT (SCMT) FOR EUROPEAN ANSPS

Kathryn Mearns
Industrial Psychology Research Centre
University of Aberdeen
Aberdeen, UK
k.mearns@abdn.ac.uk

Barry Kirwan
EUROCONTROL Experimental Centre
Bretigny
Paris, France
barry.kirwan@eurocontrol.int

Richard J. Kennedy
Boeing R&T Europe
Cañada Real de Las Merinas 1-3
Madrid, Spain
richard.j.kennedy@boeing.com

***Abstract*: This paper describes the approach used to develop a Safety Culture Measurement Toolkit (SCMT) for European Air Navigation Service Providers (ANSPs). The concept of safety culture has achieved increasing currency over the past twenty years and is now applied extensively in a number of high reliability industries including nuclear, oil and gas, manufacturing and rail, as well as the medical sector. Following a review of the safety culture literature from 2001-2005, a thematic model of safety culture was developed and items were generated from interviews with ATM personnel to reflect these themes. A questionnaire was developed, validated by ATM safety managers and deployed in eight geographically dispersed ANSPs across Europe. The questionnaire provided a 'snapshot' of the state of safety culture within the ANSP and the second phase of the SCMT involved feedback of results in workshops with ANSP personnel to reflect on the responses and determine 'why' the state of safety was perceived the way it was and how safety could be improved within the organization. Such approaches are now being applied across Europe in an effort to raise the level of safety culture in European ANSPs prior to the next 'generation' of ATM called SESAR, whose implementation phase begins 2013. This toolkit development is therefore also being strongly coordinated with FAA, who have similar safety culture ambitions for the US ATM system and its transformation via NextGen, and CANSO whose ambitions are for global improvement of safety culture in its ANSP members.**

Keywords; Safety Culture, ATM, Techniques, Validation

I. INTRODUCTION

Accident rates and perceptions of the travelling public reflect the fact that ATM has a high level of safety performance, however like most industries ATM has to constantly cope with commercial pressures and is also facing fundamental changes in the provision of ATM services. In particular, to cope with increasing demand in the next decade, two major programmes of improvement, SESAR in Europe and NextGen in the US, will transform air traffic services. Any change in an industry can challenge safety, and the scale of the changes proposed will mean that the current exemplary safety record enjoyed by ATM in general could be strained. A key question in this competitive and changing environment is therefore 'How does the industry remain safe and continue to deliver its hitherto excellent safety performance'. How will ATM personnel and systems cope with these pressures and changes, and what are the crucial ingredients of effective safety management?

Most safety critical industries have developed safety management systems (SMS) to control risks within their organizations. The SMS is a formal documented system outlining policies, procedures, rules, regulations and other types of control processes for managing risk. However, it is only as good as the people who prepare it and reinforce its principles throughout the entire organization. Employees' attitudes, perceptions, beliefs, awareness and motivation for safety can all have an impact on how well the system functions and the term 'safety culture' reflects the commitment of personnel to safety at all levels of the organization as measured through a number of related themes.

Therefore, whilst there are major efforts underway to carry out extensive safety cases for SESAR and NextGen to prepare for all potential safety threats, inevitably – and particularly with such system-wide changes occurring in parallel – there will be eventualities and scenarios which will occur that could not have been foreseen, and which may fall in the 'grey areas' between the formal rules and procedures in the Safety Management System. In such situations, our passengers' and aircrews' safety may rest in the hands of controllers who need to make the right decision, the safe decision, even when there may be strong competing pressures. Making these safe decisions in pressured, uncertain situations will depend critically on the level of safety culture in ATM's network of organisations, from the Chief Executive Officer and the executive board, to the supervisors, controllers and assistants, and the engineering staff who run the ATM systems. The over-riding aim of EUROCONTROL, FAA and CANSO's safety culture programmes is to ensure that the ATM industry is permeated with a consistently high level of safety culture, and that this high level of safety culture is in place before the realities of the transformations planned by SESAR and NextGen get underway.

II. DEFINING SAFETY CULTURE

'Safety Culture has the definitional precision of a cloud'
Professor James Reason

According to the Advisory Committee for Safety on Nuclear Installations the safety culture of an organization "is the product of individual and group values, attitudes, competencies and patterns of behaviour that determine commitment to, and the style and proficiency of, an organisation's health and safety management." [10]. The concept of safety culture has been applied in the nuclear [11] and [12], offshore oil and gas [13] [14], manufacturing [2] and railway maintenance industries [5], amongst others, however, little is know about whether the concept would apply to ATM and if so, what would be meant by the term in this industry (although see [3] and [4] for some consideration of this).

A more pithy definition of safety culture is 'the way safety is done around here', which is a neutral definition (safety culture could be positive or negative). Trademarks of safety culture are erring on the side of safety, and not being reluctant to speak up and act for safety. The precise elements of safety culture, as derived from scientific study of ATM organisations, will be described later in this paper.

It will be clear to anyone who has worked in ATM that there is already, and always has been, a strong and central emphasis on safety – in fact most definitions of ATM include the word safety (e.g. 'the safe provision of air navigation services'). ATM has always therefore had at the very least, an implicit and strong safety culture. However, the increased pressure on this industry, and the changes set to occur in the near future, call for a more explicit approach to bolster the safety culture and ensure it is not eroded. Furthermore, although other industries have evolved mature approaches to safety culture measurement and enhancement, particularly the nuclear power industry where the concept originated in the aftermath of the Chernobyl nuclear accident in 1986 in the Ukraine, ATM is a unique industry and deserves its own tailored approach.

For these reasons, EUROCONTROL in 2004 began a scientific study of Safety Culture in ATM. However, if there are ambitions to improve safety culture in many inter-related organisations, it is essential that the underlying approach has integrity and validity, so that we know we are really making improvements and do not send organisations 'down the wrong track'. Therefore, the first four years (2004-7) of the current study programme focused on development of a robust approach for any ANSP, one which would have 'face validity' with controllers and chief executives alike, and which would simultaneously satisfy the 'academics', via statistical evidence of the method's 'construct validity' and other methodological criteria applied to such methods.

The fact that eight ANSPs have already used the approach with success, to the point that it has been used to refine their corporate safety strategies, and that two ANSPs are about to embark on a second measurement and improvement cycle, suggests that the method's 'face validity' is already high. This paper therefore focuses on the somewhat more technical aspect, as to whether the approach has scientific integrity and validity. The validation process is an ongoing journey, since finding one method which fits the whole gamut of European national cultures is akin to the search for the Holy Grail, but sufficient evidence now exists to show a reasonable level of integrity and validity of the method. The rest of this paper therefore describes the approach and the evidence that this whole venture is on the right track.

III. EUROCONTROL SAFETY CULTURE TOOLKIT

The current project was set up in 2004 to investigate the following research questions:

1) What are the key themes / constituents of safety culture?
2) Do these themes apply to ATM?
3) If so, can they be measured in a meaningful way?

A. Method – Developing the Measures

There are different models of safety culture, some of which reflect a maturity dimension (e.g. [15] and [19]) while others reflect the various components believed to constitute safety culture [16] and [17]. An excellent review in [7] has summarised the safety culture literature prior to year 2000 and therefore the first phase of our study was to conduct a review of the safety culture literature from 2001-2005. This review outlined the thematic coverage of safety culture and also the methods used to investigate the construct. Emerging themes included 'How safety is prioritized in the organization', i.e. management/controller commitment, resources for safety, responsibility for safety'; 'How people are involved in safety', i.e. involving air traffic controllers, management involvement and teaming for safety' and 'How the organization identifies and disseminates lessons learned about safety', i.e. reporting incidents, communicating problems, learning from incidents, blame & error tolerance, communication about system or procedure changes, trust within an organisation, real working practices and regulator effectiveness. The most popular measurement techniques appeared to be questionnaires, interviews and focus groups, but decision tasks and role-playing scenarios [20] also featured. We decided to develop a questionnaire measure plus interviews and focus groups with the option of adding decision tasks and role-playing at a later stage in the development process.

We believed that it was important to investigate whether the identified themes were applicable to ATM services. We therefore designed a semi-structured interview schedule to investigate emerging themes with a cross-section of ANSP personnel (i.e. Air traffic controllers (ATCOs), Maintenance technicians, Engineers and Managers) across four ANSPs from northern, southern, eastern and western Europe. A total of fifty-two personnel were interviewed over a four month period at the end of 2005. One researcher conducted the interviews while the other took extensive notes. The interview was also recorded if the interviewee permitted it. The transcripts from these interviews were coded by the research team and statements reflecting the theoretical themes were identified and extracted. Five main themes containing a total of 18 elements emerged from the process as reflected in Table 1.

TABLE 1: INITIAL SAFETY CULTURE ELEMENTS AND SUB-ELEMENTS

Safety Culture Elements	Safety Culture Sub-Elements
1. Safety Management Commitment	1. Priority of safety
	2. Responsibility for safety
	3. Resources for safety
2. Trust in Organisation Safety Competence	4. Trust in Safety Process
	5. Regulatory effectiveness
3. Involvement in Safety	6. Communication about changes
	7. Communicating problems (speaking up)
	8. Involvement of ATCOs in safety
	9. Management involvement in safety
	10. Working with contractors
4. ATCO Safety Competence	11. ATCO competence
	12. Team effectiveness
	13. Training for safety in ATC
	14. Procedures & working practices
5. A Just, Reporting & Learning Culture	15. Error tolerance, blame & punishment
	16. Rewards, incentives & performance appraisal
	17. Reporting (and investigating) incidents
	18. Learning from incidents

A group of European ANSP safety managers attending a workshop in Madrid was asked to assess 80 items covering these themes by competing a questionnaire containing the statements and also rating each statement according to the following criteria: 'Must-have'; 'OK'; 'Prefer not to use'; 'Remove'. In addition, the managers were asked to give general feedback about the questionnaire regarding the following:

• Are statements realistic? Could this happen in an ANSP?

• Are statements getting at the right issues?

• Are they aimed "too high" (good example) i.e. an unreachable aim?

• Are they aimed "too low" (poor example), i.e. no ANSPs have a safety culture this poor?

• Are statements relevant for all staff, i.e. Controllers, supervisors, technicians and managers?

The feedback from the managers indicated that the face and content validity of the questionnaire was good as most of the items made sense and those that did not were discarded. However, the length of the questionnaire was criticised, with suggestions of 50 question items being a more acceptable questionnaire length for the average respondent. Following a further review by the research team, some of the items were re-worded to make them clearer and more easily understood by respondents. We also attempted to identify statements within each theme that could be considered 'enablers' of safety culture, i.e. positive, affirmative statements and also statement that were more negative and could be considered 'disablers'. In addition, four other items were added to make sure that all issues were adequately covered. The themes were as follows:

1. *Commitment to Safety*
2. *Resources for Safety*
3. *Responsibility for Safety*
4. *Involving ATCOs in Safety*
5. *Management Involvement in Safety*
6. *Teaming for Safety*
7. *Reporting Incidents/Communicating Problems*
8. *Learning from Incidents*
9. *Blame & Error Tolerance/Discipline and Punishment*
10. *Communication about Procedural/System Changes*
11. *Trust within the organization*
12. *Real Working Practices*
13. *Regulatory effectiveness*

The final questionnaire consisted of 59 items, evaluated on a 5-point Likert-type scale ranging from 1 'strongly disagree' to 5 'strongly agree'. The resulting Safety Culture Measurement Toolkit (SCMT) consisting of the questionnaire and follow-up workshops was administered to one ANSP in late 2006 and further tested at three other European ANSPs in 2007.

B. Method - Appling the SCMT in four European ANSPs

(i) Questionnaire Distribution & Representation Criterion

The method for applying the SCMT involved two stages. Part 1 entailed the distribution of the 59 item Safety Culture Questionnaire to gain a snapshot of how ANSP personnel perceived safety within their organisation. The questionnaires, to be completed anonymously, were delivered to representative samples of staff in each ANSP (including ATCOs, assistants [ATSAs], maintenance support, supervisors and managers). In order to achieve an adequate response rate (more than 30% of the target population), a "champion" within each ANSP helped to promote, distribute and encourage completion of the questionnaires. The target response rate of over 30% was reached in three of the four ANSPs, with the highest being ANSP B: n= 310 responses (70%). The criterion level was not reached in ANSP C (n=50), although we still attempted to insights into safety culture within that organisation using this sample.

(ii) Questionnaire Analysis

The completed questionnaires were returned to the project team for analysis. Data were entered into a spreadsheet, and sample means, ranges and standard deviations were calculated to identify trends of favourable and unfavourable responses. Potential problem areas were also identified, defined as any questionnaire items that were given an unfavourable response

by 25% or more respondents. On the basis of these data, the project team identified a number of safety culture "themes" for each ANSP, based on the framework developed from the literature review, the interviews with ANSP staff and the safety managers' workshop, which were to be the focus of the second part of SCMT application.

(iii) ***Workshops***

Part 2 of the process involved conducting workshops or interviews with ANSP staff, designed to explore the areas of interest identified from the analysis of questionnaire responses. In ANSP A, nine interview sessions were conducted involving 11 personnel from a variety of roles within the organisation. These interviews were carried out with the aims of validating the results of the questionnaire, gaining a deeper understanding of the issues raised, and identifying examples of "good practice" relating to the safety culture themes arising from that ANSP's questionnaire responses. In ANSPs B, C and D, focused workshops were conducted rather than interviews. This was because although the interviews were useful, they were methodologically less robust than focus groups with multiple personnel. In interviews, essentially people can say what they like 'unchallenged' and without any corroborative evidence. In facilitated focus groups, particularly with people from different work areas, the collective insights tend to be sharper and more *robust.*

The aims of these small group workshops were to gain deeper insight into areas of interest identified from questionnaire responses, to provide insight into why these issues were thought to be important and to identify potential solutions. In these three ANSPs, four or five workshops were conducted, each involving five employees representing all levels of the organisation. The workshops began with the project team explaining the concept of safety culture, its importance, the SCMT approach and the aims of the workshop. The safety culture themes arising from the questionnaires were then outlined and discussed, with participants providing their thoughts on each issue. The items were separated into main themes and prioritised in terms of importance, and potential solutions were then discussed.

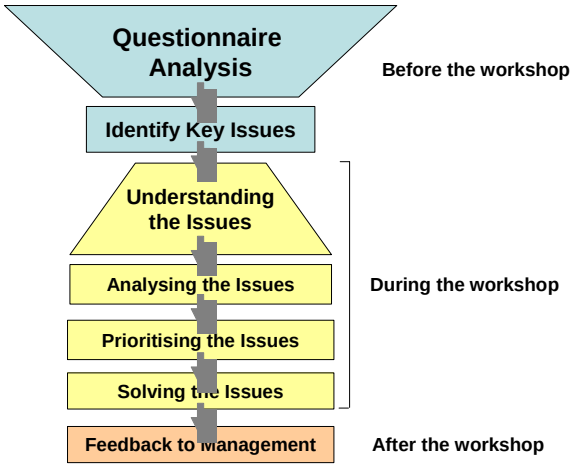
It should be noted that generally the workshops were dynamic and generative of ideas, suggesting a strong motivation to improve safety culture.

(iv) ***Reporting***

Finally, each ANSP received a report combining the results of the questionnaire analysis and the workshops. Recommendations included items which could be considered 'quick wins' (e.g. clearer communications on the priority of safety in organisational [corporate] media), tactical short term measures (e.g. ensuring better consistency in the approach of supervisors to discipline following controller mistakes), to medium and longer term strategic measures (e.g. re-instatement of Team Resource Management [TRM] practices, and 'outreach' safety culture campaigns to carry the safety culture message to every outpost of the more geographically disparate ANSPs). It should be emphasised that the final decisions on what actions to implement rested with the ANSPs themselves, in particular the executive management board. Lastly, as well as the formal report to Management, there is also a presentation of the results and recommendations to the general workforce who participated in the survey. The decision whether and how to disseminate the report or excerpts from it then rests with the Management.

The methodology adopted for the full SCMT process is illustrated in Figure 1. Application of the process takes between four and six months. It can be repeated after 2-3 years, to determine if safety culture has improved (the first two of the four ANSPs surveyed will launch their 'second cycle' of safety culture surveys in 2009).

FIGURE 1: SCMT ANALYSIS PROCESS



C. *Validity of the Measurement Approach*

A key aspect of this project was to determine the validity of the SCMT. There are various types of validity. Face validity refers to the extent to which the method and measures appear to be suitable for the task in hand. Content validity refers to whether the method or measure represents the relevant area of interest and construct validity is the appropriateness of the theory behind the method. Discriminant validity refers to the extent to which individual items measures constructs that are distinct from one another, i.e. items do not have overlapping content. Finally, predictive validity refers to whether the measures and method actually predict future events, such as improved safety performance.

It was not feasible to address predictive validity in this project since this would require a longitudinal approach to determine whether current interventions designed to improve the safety culture actually have an impact on the organization's safety performance. There are other factors, which confound evaluations of predictive validity. For example, safety culture failings may lie dormant in the system for considerable periods of time, i.e. months to years before they are realized as adverse events [17]. Similarly, improvements in safety culture may be successful in averting incidents, however, if these incidents are prevented then nothing happens and therefore the effectiveness of the actions will go unnoticed. For these reasons face, content, discriminant and construct validity were evaluated for the questionnaire phase of the SCMT and face and content validity were evaluated from the Workshop phase.

D. Evaluating the Questionnaire Measure

The stages in the implementation of the SCMT are distribution of the questionnaire to a cross-section of ANSP personnel, return of adequate data for analysis purposes, data analyses and interpretation and production of a feedback report to the ANSP. The survey is a relatively quick and easy way of collecting data from a large sample of people. In this sense it is user-friendly and requires minimum effort from the participants and furthermore, the questionnaire can be completed without any training or extensive experience. Factors that detract from a questionnaire approach are a lack of control regarding where and when the questionnaire is completed; the possibility of collaboration when completing the questionnaire (therefore responses are not independent of each other) and aspects such as social desirability (i.e. the need to present oneself or one's colleagues or the organization in a positive light). There is also the issue that poorly motivated individuals may complete the questionnaire items in a random way or without due consideration of the statements posed. Measures can be taken to alleviate these problems, for example, stressing the anonymous and confidential nature of the data collected; providing the right conditions for the questionnaire to be completed, e.g. individually, at a particular location where people are supervised throughout the process; and providing incentives to achieve high response rates from the sample.

It is proposed the examples of 'best practice' regarding how to manage questionnaire completion, should be disseminated to participating ANSPs and these should be directed at Senior Management and cascaded down to supervisors and the workforce in order to achieve their commitment and buy-in. This should lead to adequate deployment of resources to achieve the most advantageous outcome for the organization in terms of time invested in the process and the accuracy and integrity of the data collected. In many ways, the deployment of adequate resources ensures that the data collected are useful are provide the most accurate and valid insight into the safety perceptions and beliefs of ANSP personnel.

In practice, two recurrent 'reluctance factors' can be encountered: first, at the management level, a fear that the results will in some ways incriminate them and their failings (whether real or perceived), opening up a 'can of worms'. Second, the workforce may fear that this is just another questionnaire whose results will be ignored or 'white-washed'. The solution to the first is that the more responses gained, then the less tendency there is to have 'extreme' responses dominating the resulting safety culture profile. Additionally, the workshops represent a way of 'moderating' questionnaire results, such that negative findings in safety culture terms can at least be challenged and probed. If serious failings are found, then management need to know them and act on them. The second fear is best countered by the fact that the questionnaire is run independently by experts outside the ANSP. At survey 'launch' events, wherein presentations are made at general meetings at ANSP HQs and regional centres, it is stressed that the questionnaire offers people a sincere chance to have their say on safety and to be heard. As an extraordinary measure, 'hotline' phone numbers and email addresses are given to any participant, in case they wish to raise sensitive issues direct to the survey leaders. In both cases of such 'reluctance factors', the central counter-argument is the same: ANSP personnel don't want accidents, the best solution is to have a clear rather than a 'rosy' risk picture, and so the best approach is to gain a strong and representative response rate to the questionnaire and balanced participation in the ensuing workshops. This is indeed the conclusion of a number of ANSPs who have now been through this process, and what their safety directors and even their CEOs tell other ANSPs.

With respect to the validity of the questionnaire measure, it is apparent from feedback from ANSP Managers at the Madrid workshops and also comments from personnel who have completed the questionnaire that the statements have face validity, i.e. most of the items make sense to the respondents and cover issues considered to be of importance for good safety performance. That said, certain items were identified by respondents as not being fully understood, but these items tended to vary from country to country. One of the main issues was that the statements were perceived to be too ATCO-orientated or too technical in nature and were either 'not applicable to' or 'not understood by' other disciplines within the ANSP. One of the key recommendations from this observation (corroborated by the factor analytical work) is that a central core of safety culture items that are relevant across disciplines and individuals within ANSPs are utilized for organizational safety culture appraisal, but then respondents are specifically routed to sections dealing with discipline-specific safety culture issues once the central items have been completed. For example, engineering and maintenance staff should be routed to a section that is relevant for their activities and air-traffic controllers, their supervisors and assistants should respond to items that are designed with their working practices in mind. The current version of the questionnaire in use has four sections: a general section which everyone completes, followed by three mutually exclusive sections to be completed by management, engineers, and the operational controller workforce respectively.

The items used in the questionnaire were originally derived from a theoretical framework based on a literature review which identified the factors believed to constitute safety culture. The review of the safety culture literature from 2001-2005 indicated a number of consistent themes. These were 'How safety is prioritized in the organization', 'How people are involved in safety' and 'How the organization identifies and disseminates lessons learned about safety'. Since our ideas about safety culture are theoretically grounded, the underlying constructs (or themes) being measured by the questionnaire should have construct validity, however this assumption needs to be tested using statistical modeling. We also need to use the statistical modelling to determine the discriminant validity of the items, i.e. do they measure the construct of interest clearly and distinctly. Finally, the content validity of the items was tested by conducting interviews with ANSP personnel, to check whether these themes made sense in relation to their organization.

IV. STATISTICAL VALIDATION

A. *Iteration 1: Exploratory Factor Analysis*

Using the Statistical Package for the Social Sciences (SPSS 15), Exploratory Factor Analysis (EFA) was conducted on two of the datasets; ANSP A and ANSP B. EFA is a statistical technique that is used to understand the structure of a set of variables (in this case questionnaire items), and to reduce a data set to a more manageable size while retaining as much of the original information as possible. On the basis of an EFA it should be possible to distinguish an underlying structure to the data with specific items loading (i.e. correlating with one another) on specific dimensions, e.g. one should be able to identify a dimension measuring 'Involvement in Safety' and another measuring 'A Just and Learning Culture'. One would anticipate that the items that are designed to measure these concepts would correlate highly with these dimensions. In order to run EFA a certain sample size is recommended. [18] recommend at least 300 cases and also [6] notes 1000 as an excellent sample size, 300 as good and 100 as poor. Given that only the sample sizes for ANSP A and ANSP B exceeded the 'poor' criterion of 100, EFA was only conducted on these datasets. Unfortunately, the underlying factor structures that emerged for each sample were difficult to interpret. For example, the ANSP A dataset initially indicated a total of 18 factors but items that correlated with each other statistically did not necessarily make sense from a thematic perspective. We tried to force alternative analyses with solutions ranging from 2-13 factors but none of the solutions provided easily interpretable dimensions. Finally, the best fit was found with an 8-factor model but still the items that grouped together statistically did not reflect a single specific factor. For example, the 'Reporting' factor actually contained items that we would expect to appear under 'Trust' and 'Commitment'.

For the ANSP B dataset, only one factor emerged with our initial EFA, i.e. an overall safety culture dimension and even when 13 and 8 factor solutions were imposed by the programme this single factor persisted. In other words it was not possible to distinguish separate underlying safety culture 'themes' such as 'Reporting'; 'Teaming for Safety'; 'Learning from Incidents' and 'Commitment to Safety'. This was disappointing since the ANSP B dataset with 310 respondents was the most appropriate for EFA. These analyses seem to indicate that it is possible to achieve a single unitary factor which measures 'Safety Culture', however we were not able to distinguish the underlying constructs that we believe constitute safety culture such as 'Commitment'; 'Resources for Safety'; 'Communication' and so on.

B. *Iteration 2: Confirmatory Factor Analysis*

Confirmatory Factor Analysis (CFA) is a more sophisticated technique and is used in later stages of the research process to corroborate theory about the underlying relationship between latent variables (Tabachnick and Fidell, 2001). For the purposes of this study we had a prior, strong theoretical basis for expecting specific relationships between variables, based on our literature review and our initial assumptions when the research programme began, supported by the Safety Managers at the Madrid workshop, as well as by questionnaire feedback where almost no additional factors or issues were suggested by any respondents in response to an open question at the end of the questionnaire.

The most common form of CFA is structural equation modelling (SEM). In this type of analysis, a theoretical model is set up and the data are analysed in order to determine how well they 'fit' the theoretical model. Various parameters are used to determine this 'Goodness of fit' including the Goodness of Fit Index (GFI) and the Comparative Fit Index (CFI). Values of .90 or better on these indices are indicative of a good-fitting model. Another measure, the root mean square error of approximation (RMSEA) estimates the lack of fit in a model as compared to a perfect fit. Values of .06 or less indicate a good-fitting model. Values larger than .10 indicate a poor fitting model. A final measure of fit is Chi2 (X2) where a good fit is indicated by a non-significant X2. A non-significant Chi2 would indicate a consistent model across the dataset. In the current study, it is relevant to examine the ANSP datasets using CFA since we are testing a theory about the underlying factors that constitute safety culture, however, certain caveats need to be observed, for example that the sample size should be about 200 for an adequate model to be generated. As a result of this caveat we only report the CFA conducted on the ANSP B dataset.

Since there was little evidence from the EFA that our hypothesised structure of 5 main dimensions (or themes) with 13 subscales (or sub-themes existed, we decided to return to the 3 main order factors we identified from the literature review and tried to fit 7 or 8 sub-themes under these higher order factors. We had good evidence of the validity of these main order factors and sub-themes since they had been approved at the managers' workshop in Madrid. In other words, domain experts believed that these concepts helped describe the essence of safety culture. These higher order factors were stipulated as 'How safety is prioritized in the organization' (Prioritization); 'How are people involved in safety' (Involvement), and 'How the does the organization identify and disseminate lessons learned about safety' (Learning) and the sub-themes were 'Commitment'; 'Involvement'; 'Priority of Resources'; 'Reporting'; Learning; 'Teaming'; 'Trust' and 'Communication'. After a number of iterations we achieved a good model fit for ANSP B (GFI = .898, CFI = .931, RMSEA = .034) with 30 items mapping onto four sub-factors, which in turn mapped onto three higher order factors. Each sub-factor had four to six questionnaire items correlated with it. The higher order factors were, 'How we prioritize safety (the sub-factors Commitment and Responsibility were combined in this higher order factor'; 'How we are involved in safety (Involvement and Teamwork) and 'How we Learn (Reporting, Communication and Trust), which provides us with a basis for further work. This was an important finding, suggesting first that Safety Culture itself is a valid construct, and second that it can be meaningfully de-constructed to lower levels, such that questions can be formed to determine safety culture in a scientifically meaningful way, one that is also meaningful to respondents. It means we are measuring what we hope to measure. The necessary caveat is that this result needs to be corroborated by further large datasets. One recent survey has

indeed produced a dataset in excess of 500 responses, and so will be analysed statistically using CFA to see if the same 'model' is supported.

C. *Validation of the Workshops*

The second phase of the validation process of the SCMT addressed the Workshops, which were used as a forum to understand the processes underlying the ANSP's safety culture and identify ways in which the safety culture could be improved within the ANSP. It was decided that the themes to be discussed at the workshops should be based on the proportion of statements that had more than 25% of respondents (one in four) answering the statements in a negative way. The statements were than categorized according to the themes it was felt they reflected best, e.g. 'Priority of Safety'; 'Involvement in Safety' and 'Learning' and so on. There are no quantitative techniques that would help determine the validity of the workshops from a psychometric perspective, however, a qualitative appraisal can be made bearing in mind the criteria outlined earlier. The final column Table 2 indicates whether the workshops met the requirements of these qualitative criteria.

TABLE 2: TESTING CRITERIA FOR WORKSHOP PHASE OF SCMT

Definition	Qualitative Measures	Workshop Performance
Validity - Approach measures what it purports to measure.	Face validity tests such as feedback from domain experts (e.g. ANSP Safety Managers).	The workshops appear to have good face validity as the themes covered and the discussions about specific items seemed to make sense to participants.
Comparative Validity – the SCMT should perform similarly to other comparable safety culture survey approaches	Review against other safety climate / culture approaches being deployed in European ATM.	We had the opportunity to evaluate the SCMT against another safety culture evaluation in one of the ANSPs, applied independently by a University. The two approaches led to very similar results and highlighted similar issues of concern.
Consistency - Approach produces consistent results when applied by different assessors.	Involvement of different members of the Project Team in the various ANSP trials. Compare outputs to judge whether consistent.	Different Project Team members took part in the various workshops and although the outputs from the workshops varied according to ANSP there was a consistency in the process used..
Usability - Approach easy to use without requiring extensive experience and training.	Feedback from ANSPs who follow the process without the support of the Project Team. Define training requirements for someone to be able to lead a study.	As far as we know none of the ANSPs have followed the workshop process without the support of the Project Team. Specific training may be required for people to lead the study and it may be better for a group of people outside the organization to conduct the workshops.
Usefulness - Approach provides outputs that are compatible with other tools and techniques for safety management..	Evaluate how the SCMT fits into the suite of assessment tools that can be used for safety management and culture assessment	A number of the recommendations arising from SCMT application impact areas of the Safety Management System (SMS) as well as informing the broader Strategic Safety Plans of ANSPs.
Insight - Approach provides additional information and insights beyond the approaches currently deployed	Degree to which the SCMT results in new initiatives by the organisation.	In three of the four ANSPs the SCMT application led to new initiatives to improve safety culture, in two cases leading to a major safety strategy shift to improve safety culture. The one ANSP where this did not happen was also the one where the response rate was well below the standard criterion level.
Resources - Approach produces outputs that are commensurate or exceed that expected with the resources deployed.	Consideration as value for money by the ANSP management	In December 2008 at a special CEO-level Safety Culture Conference in Rome, several ANSPs went on record stating that the results were well worth the resources expended, and that they now had a richer and more accurate 'risk picture' than before.
Auditable - Approach produces outputs that are explicit, traceable and open to scrutiny.	3rd party peer review of project and ANSP reports	The initial results of independent audit and review of the SCMT were generally favorable.

V. DISCUSSION AND CONCLUSIONS

A. *Validity*

This current study reports on the design of a Safety Culture Measurement Toolkit for ATM across European Member States. Based on an up-to-date review of the safety culture literature, a model of what constitutes safety culture was defined and appropriate measurement techniques were developed. These included a questionnaire and facilitated workshops. We have focused very much on ascertaining the validity of our approach. In the questionnaire phase, face and content validity appeared to be good, however we continue to collect data to verify that the safety culture model elucidated based on the largest ANSP sample fits across multiple European ANSPs. It is of course possible that there is no 'one-size-fits-all' structure, due to underlying fundamental differences in national cultures; this will be determined as new and sufficiently large datasets are analysed. Indeed, safety may be a universal value but it may be construed in different ways by different people. In practice this may mean that there is a similarity in the overall safety culture framework and its elements, but that these elements are arranged differently in

different cultures. This will be reported on in future publications as the analysis proceeds.

We could not assess the validity of the workshops using statistical models, but qualitative feedback suggested the workshops had good face and content validity. The workshops met the qualitative criteria of consistency, usability, insight, usefulness, auditability, and 'value for money'.

B. (ii) Questionnaire Refinement

Our models indicated that some questionnaire items were correlated and therefore had overlapping content and poor discriminant validity. These items have since been refined and clarified to reflect a single construct. Respondents also seemed to be confused by negative wording and there is recent evidence to suggest this is a universal problem in questionnaire design (REFS). Translation may have further obfuscated the meaning of items (one way to counteract this is to 'back-translate' the items into English). A further recommendation would be to create positively worded items throughout the questionnaire for clarity and consistency and to back-translate questionnaires from ANSPs where English is not the native language. The questionnaire is currently being applied (for the first time) to a native English-speaking country, so the precision and meaningfulness of the questionnaire items can be assessed without the usual translation or foreign speaker confounding factors.

To achieve more consistency in our modelling process we would require samples of at least 200 people; recently, two such datasets have been gained, and will be analysed using EFA and CFA techniques. Furthermore, two or three large ANSPs will be surveyed in the 2009-10 timeframe with an adapted version of the questionnaire that contains items with improved construct and discriminant validity. The latest version of the questionnaire contains a core set of items which can be answered by all ANSP personnel, irrespective of profession or occupation, followed by three occupation-specific sections (management; engineering; ATCOs). Early feedback suggests this is an improvement.

C. Is Safety Culture an Issue?

Although some aspects regarding the validity of the questionnaire still need to be addressed, the workshops have proved to be successful in achieving their objective of trying to understand the processes behind the creation of the current safety culture within the organization. For example, personnel identified change processes and lack of communication about change as risk factors, and the perception that throughput of traffic takes precedence over safety issues seemed to be endemic in the ATM organizations we surveyed (this is also the case for the oil and gas industry). It is important to note that safety is not disregarded, but safety issues are expected to be addressed while throughput is maintained, in some cases even if the system is working in a degraded mode without some of its usual functions. The resolution of this issue continues to challenge high reliability organizations, but may be especially difficult for ATM, since the controller is making decisions in real-time, under pressure, and often in situations involving uncertainty, generally against a gradual increase in traffic year on year.

Staffing levels, shift rotas and communication between different parts of the organization were also identified as issues and solutions, as well as recruitment of more ATCOs and team resource management training were discussed in the workshops. A major issue arising in most ANSPs surveyed is the continued assurance of safety during periods of change, such as moving to a new Centre with a new ATC system, human-machine interface, and procedures. This is of particular significance given the approaching implementation phases of SESAR and NextGen in the medium term future. It reinforces the importance of ensuring a high level of safety culture across European and other ANSPs before anticipated system-wide changes commence. In this way, a strong Safety Culture, together with a strong SMS, will provide a robust 'safety backbone' to see ATM safely through its transition to the next generation of ATM.

VI. REFERENCES

[1] Cooper, M. D. 2000. Towards a model of safety culture. Safety Science, 36, 111- 136.

[2] Cox, S., Tomas, J. M., Cheyne, A., & Oliver, A. (1998). Safety culture: The prediction of commitment to safety in the manufacturing industry. British Journal of Management, 9, S3-S11.

[3] Ek, A., & Arvidsson, M. (2002). Safety culture in the Swedish air navigation services. Paper presented at the European Academy of Occupational Health Psychology, Vienna, 4-6 December.

[4] Ek, A., Arvidsson, M., Akselsson, R., Johansson, C., & Josefsson, B. (2003). Safety culture in air traffic management: Air traffic control. Paper presented at the 5th USA/Europe ATM 2003 R&D Seminar, 23-27th June, Budapest, Hungary.

[5] Farrington-Darby, T., Pickup, L., & Wilson, J. R. (2005). Safety culture in railway maintenance. Safety Science, 43(1), 39-60.

[6] Field, A. 2003. Discovering Statistics using SPSS for Windows. London: Sage Publications.

[7] Guldenmund, F. W. (2000). The nature of safety culture: a review of theory and research. Safety Science, 34(1-3), 215-257.

[8] Helmreich, R.L. & Merrit, A.C. 1998. Culture at work in aviation and medicine: national, organisational and professional influences. Aldershot: Ashgate.

[9] Hofstede, G. (1994) Value survey module 1994 manual. Netherlands; IRIC.

[10] Health and Safety Commission (1993). Organising for Safety: Third report of the Human Factors Study Group of ACSNI (Advisory Committee on the Safety of Nuclear Installations). Sudbury, HSE Books.

[11] Lee, T. 1998. Assessment of safety culture at a nuclear reprocessing plant. Work and Stress, 12, 217- 237.

[12] Lee, T . & Harrison, K. 2000. Assessing safety culture in nuclear power stations. Safety Science, 34, 61 – 97.

[13] Mearns, K., Whitaker, S. M., & Flin, R. (2003). Safety climate, safety management practice and safety performance in offshore environments. Safety Science, 41(8), 641-680.

[14] Mearns, K., Whitaker, S. M., & Flin, R. (2001). Benchmarking safety climate in hazardous environments: a longitudinal, inter-organizational approach. Risk Analysis, 12(4), 771-786.

[15] Parker, D., Lawrie, M. & Hudson, P. 2005. A framework for understanding the development of organisational safety culture. Safety Science, 44, 551-562.

[16] Reason, J. (1997). Managing the risks of organisational accidents. Aldershot: Ashgate.

[17] Reason, J. (1998). Achieving a safe culture: Theory and practice. Work & Stress, 12(3), 293-306.

[18] Tabachnick, B.G. & Fidell, L.S. (2001) Using Multivariate Statistics. 4th Edition, Boston: Allyn and Bacon.

[19] Westrum, R. (2004). A typology of organisational cultures. Quality, Safety & Healthcare, 13, 22-27.

[20] Zohar, D., & Luria, G. (2004). Climate as a Social-Cognitive Construction of Supervisory Safety Practices: Scripts as Proxy of Behavior Patterns. Journal of Applied Psychology, 89(2), 322-333.

AUTHOR BIOGRAPHY

Dr. Kathryn Mearns is a Senior Lecturer and Researcher in the Industrial Psychology Centre at the University of Aberdeen. In addition to ATM, she is currently conducting safety culture R&D programmes in the medical and offshore industries. She holds Bachelors and PhD Degrees in Psychology.

Dr. Barry Kirwan is the Head of Safety Improvement Research at EUROCONTROL. He has worked for many years in a number of different high reliability organizations including nuclear, offshore and Air Traffic Management. He holds a Bachelors Degree in Psychology, a Masters Degree in Human Factors and a PhD Degree in Human Reliability Assessment.

Dr. Richard J. Kennedy is the Manager of the New Programs and Safety Group at Boeing Research & Technology – Europe (BR&TE). His experience covers nuclear, rail, ATM and flight deck domains. He holds a Bachelors Degree in Psychology, a Masters Degree in Human Factors, a Master of Business Administration (MBA) and PhD Degree in Manufacturing and Mechanical Engineering. He is also Chartered Engineer (CEng).

DISCLAIMER