Enhancing Performance in High Risk Environments –
Recommendations for the use of Behavioural Markers

Note: Content with permission of Gottlieb Daimler and Karl Benz Foundation Kolleg Group Interaction in High Risk Environments (GIHRE).

Intentionally Left Blank

GOTTLIEB DAIMLER-
UND
KARL BENZ-STIFTUNG

GIHRE

Group Interaction in High Risk Environments

Enhancing Performance in High Risk Environments

Recommendations for the use of Behavioural Markers

ENHANCING PERFORMANCE IN HIGH RISK ENVIRONMENTS:

Recommendations for the use of Behavioural Markers

GIHRE

Behavioural Markers Workshop

Sponsored by the
Gottlieb Daimler and Karl Benz Foundation
Kolleg Group Interaction in
High Risk Environments (GIHRE)

B. Klampfer, R. Flin, R. L. Helmreich, R. Häusler, B. Sexton, G. Fletcher,
P. Field, S. Staender, K. Lauche, P. Dieckmann, A. Amacher

GROUP INTERACTION IN HIGH RISK ENVIRONMENTS

GIHRE-Aviation: Swiss Federal Institute of Technology (ETH) Zurich
University of Bern
University of Texas at Austin Human Factors Research Project
Industrial Psychology Group, University of Aberdeen

Swissair Training Centre
Zurich
5 - 6 July 2001

Enhancing Performance in High Risk Environments

Recommendations for the use of Behavioural Markers

CONTENTS

INTRODUCTION

ENHANCING PERFORMANCE IN HIGH RISK ENVIRONMENTS:

Recommendations for the use of Behavioural Markers

There is general agreement regarding the importance of interpersonal behaviours in technological environments and the need for training these behaviours (sometimes called non-technical skills or behavioural markers) to supplement technical instruction. Crew Resource Management (CRM) training was instituted to accomplish this in aviation and is now mandated worldwide. Behavioural markers have been used in aviation as exemplars in training and in research for assessment of CRM practices and the impact of training. However, there appears to be growing concern with the assessment of behaviours taught in CRM and increased regulatory pressure for their formal evaluation.

The assessment of CRM performance in aviation is based on consensus regarding CRM skills and associated behavioural markers, which serve as indicators of how effectively CRM concepts are enacted, whether in simulation or actual flight. Behavioural marker systems are now being developed for performance measurement in a range of organisational settings, especially in high reliability industries such as nuclear power, rail and maritime transport, and medicine. The reasons for using behavioural marker systems differ and so do the existing behavioural marker systems, even if core concepts are very similar. There seems to be some confusion as to what exactly behavioural markers should be used for and how they can contribute to better performance in operational environments.

When Swissair faced these issues, Capt. Werner Naef, Head of the Human Factors Training Department took the opportunity to establish a simulator study aimed at validating and investigating CRM behavioural markers in high workload situations. The resulting project, GIHRE-aviation (see Appendix 1), was launched within the framework of the Kolleg Group Interaction in High Risk Environments (GIHRE) funded by the Gottlieb Daimler and Karl Benz Foundation, which is examining teams in environments ranging from flight decks and hospital operating theatres to nuclear power plant control rooms (http://www2.rz.hu-berlin.de/GIHRE/). The University of Texas Human Factors Research Project, a participant in the GIHRE project (www.psy.utexas.edu/helmreich), developed one of the behavioural marker systems used in the GIHRE-aviation project and in Line Operations Safety Audits (LOSA: see Appendix 2). The other behavioural marker system came from the European Non-technical Skills project (NOTECHS: see Appendix 3).

A research team of industrial psychologists at the University of Aberdeen, Scotland, (www.psyc.abdn.ac.uk/serv02.htm) is studying non-technical skills of anaesthesia and nuclear power plant operator teams. In addition, they have been a partner in the European projects NOTECHS and Joint Aviation Requirements – Translation and Elaboration of Legislation (JARTEL) developing a behavioural marker system for airline pilots. Barbara Klampfer from the Zurich group and Rhona Flin from the Aberdeen team decided that it would be beneficial to hold a joint workshop to share research experiences and to discuss the development and utilisation of behavioural marker systems. The Swiss group organised a two-day meeting hosted at Swissair 5-6 July 2001. Attendees[1] and their affiliation are shown in Table 1.

Rhona Flin	University of Aberdeen
Georgina Fletcher	University of Aberdeen
Kristina Lauche	University of Aberdeen
Paul Field	British Airways & JARTEL Project
Barbara Klampfer	Swiss Federal Institute of Technology/GIHRE
Ruth Haeusler	University of Bern/GIHRE
Andrea Amacher	University of Bern/GIHRE
Robert Helmreich	University of Texas at Austin/GIHRE
J. Bryan Sexton	University of Texas at Austin/GIHRE
Sven Staender	Hospital of Maennedorf, Zurich, Switzerland
Peter Dieckman	University of Heidelberg

Table 1. Workshop attendees and their affiliations

Discussion at the workshop confirmed that there appear to be many misconceptions regarding the strengths and weaknesses of behavioural marker systems for the measurement of non-technical skills. The term “behavioural marker system” is used to refer to a taxonomy or listing of key non-technical skills associated with effective, safe job performance in a given operational job position (e.g., flight deck crew), with some decomposition of major skill areas (e.g., decision making) usually illustrated by exemplar behaviours. Taking into account fundamental concerns about the validity and quality of any system for assessing CRM behaviour, the need for a clear and simple document on behavioural marker systems – not focusing on a specific behavioural marker system, but on the general concepts and their application – was acknowledged. Furthermore, it was agreed that a critical factor in the implementation of any assessment tool is the training of the users of the system. Consequently, the specification of precise training requirements for the qualification of those who apply the system was seen as an essential need.

[1] Werner Naef, (Swissair), Gudela Grote, (ETH), and Paul O’Connor (University of Aberdeen) were invited, but unable to attend.

It was therefore decided that a major goal of the workshop should be to produce a set of general guidelines for practitioners and researchers who apply or are considering employing behavioural marker systems in training, development, and performance monitoring. It was also concluded that clarification of these issues might be helpful for regulators specifying requirements for the use of such systems.

BACKGROUND

The guidelines are presented below as a set of summary statements about the main features of behavioural marker systems. These behavioural taxonomies were first developed for research and training purposes in the aviation industry, the best-known example being the University of Texas (UT) Behavioural Markers developed by the University of Texas Human Factors Research Project. A subset of these behavioural markers is included in the LOSA (Line Operations Safety Audit) system used in aviation for non-jeopardy assessment of system performance and safety. In addition to behavioural markers, LOSA takes contextual factors (external and internal threats to safety, including environmental and operational conditions, crew experience and composition, etc.) into account and specifically quantifies threats and errors and their management. In this context, the non-technical skills measured addressed by the behavioural markers are defined as threat and error countermeasures. The development of the UT markers and the LOSA system is described in more detail in Appendix 2.

A number of airlines have developed their own behavioural marker systems for training and assessing flight crew skills (see Flin & Martin, 2001, for a review), although these systems are not available in the public domain. The European aviation regulator Joint Aviation Authorities (JAA) identified a requirement for a European behavioural marker system. Two international research projects were funded to develop a prototype non-technical skills marker system (NOTECHS project) and then to conduct an experimental and operational test of the resulting NOTECHS system (JARTEL project). These are described in Appendix 3 with details of the NOTECHS methodology.

By the late 1990s, behavioural marker systems for training and assessing non-technical skills were being developed for other workplaces requiring high levels of individual and team performance, such as nuclear power plants and hospital operating theatres. The GIHRE project encompasses a number of different investigations of group performance, one of which utilises both the NOTECHS and the University of Texas (UT) Behavioural Markers for rating pilots' performance. The NOTECHS

INTRODUCTION

and UT systems were designed for different purposes, but they essentially measure a similar set of component behaviours as shown in Appendix 4.

This document is only intended to convey basic information about the derivation and operational use of behavioural marker systems. These are set out in the form of 17 Frequently Asked Questions and a conclusion. It is not intended to provide guidance on the development of such systems or their use for research purposes. A bibliography is provided, which offers suggestions for further reading on the subject.

Acknowledgements

We would like to acknowledge the constructive feedback provided by the following individuals who read the draft document: Steve Belton, Capt. Colin Budenberg, Margaret Crichton, Dr Ronnie Glavin, Dr Jurgen Hoermann, Capt. Mike Lodge, Capt. Dan Maurino, Dr Rona Patey and John A. Wilhelm. Research support for the development of the University of Texas Behavioural Markers was provided by Federal Aviation Administration Grant 99-G-004.

FREQUENTLY ASKED QUESTIONS ABOUT BEHAVIOURAL MARKER SYSTEMS

BEHAVIOURAL MARKERS

1. What are behavioural markers?

- Observable, non-technical behaviours that contribute to superior or substandard performance within a work environment (for example, as contributing factors enhancing safety or in accidents and incidents in aviation)
- Observable behaviours of teams or individuals
- Usually structured into a set of categories
- The categories contain sub-components that are labelled differently in various behavioural marker systems (e.g., NOTECHS: “elements“ and “markers“ = UT (LOSA): “anchors“).

2. How are behavioural markers derived?

- From analysis of data from multiple sources regarding performance that contributes to successful and unsuccessful outcomes (e.g., accident investigation, confidential incident reporting systems, incident analysis, simulator studies, task analysis, interviews, surveys, focus groups, ethnographies)

3. What makes a good behavioural marker?

- It describes a specific, observable behaviour, not an attitude or personality trait, with clear definition (enactment of skills or knowledge is shown in behaviour).
- It has demonstrated a causal relationship to performance outcome.
  – It does not have to be present in all situations.
  – Its appropriateness depends on context.
- It uses domain specific language that reflects the operational environment.
- It employs simple phraseology.
- It describes a clear concept.

4. What are the domains of application?

Behavioural markers can be used in any domain where behaviour relating to job performance can be observed. However, they are expensive to develop and utilise given the level of training and calibration required for users. At present, they tend to be found in occupations where safety is prime and high fidelity simulators are used for training and assessment, e.g., in aviation, nuclear power generation, military settings, and, to a lesser degree, in medicine (anaesthesia and surgery), where simulation is less widely employed.

5. What are the uses of behavioural markers?

- To enable performance measurement for training and assessment, evaluation of training, safety management, and research
- To highlight positive examples of performance
- To provide a common vocabulary for training, briefing and debriefing, communication, regulation, research, and to connect different domains of safety (e.g., incident analysis and performance tracking)
- To build performance databases to identify norms and prioritise training needs
- To compare sub-groups in organisations (e.g., aircraft fleets, etc.)
- To give feedback on performance at individual, team, organisational, and system levels
- To establish co-operation between safety/quality, training, and operations

BEHAVIOURAL MARKER SYSTEMS

6. What are characteristics of good behavioural marker systems?

- Validity: in relation to performance outcome
- Reliability: inter-rater reliability, internal consistency
- Sensitivity: in relation to levels of performance
- Transparency: the observed understand the performance criteria against which they are being rated; availability of reliability and validity data
- Usability: easy to train, simple framework, easy to understand, domain appropriate language, sensitive to rater workload, easy to observe
- Can provide a focus for training goals and needs
- Baselines for performance criteria are used appropriately for experience level of ratee (i.e., ab initio vs. experienced ratees)
- Minimal overlap between components

7. What are the limitations of behavioural marker systems?

Cannot capture every aspect of performance and behaviour due to:

- Limited occurrence of some behaviours
  – These are important but infrequent behaviours, such as conflict resolution.
- Limitations of human observers – distraction, overload (e.g., in complex situations, large teams)

8. What considerations must be made when using a behavioural marker system?

- Raters require extensive training (initial and recurrent) and calibration.
- Behavioural marker systems do not transfer across domains and cultures without adaptation (e.g., western markers in eastern cultures, or from aviation to medicine).
- Behavioural marker systems need proper implementation into an organisation, and need management and workforce support.
  – phased introduction of behavioural marker systems required to build confidence and expertise in raters and ratees
- Application of the behavioural marker system must be sensitive to the stage of professional development of the individual, and to the maturity of the organisational and professional culture (e.g., whether used as a diagnostic, training, and/or assessment tool).
- Use must consider context (e.g., crew experience, workload, operating environment, operational complexity)

9. What are special considerations when using a behavioural marker system for assessment?

The use of a behavioural marker system in a formal assessment of non-technical aspects of performance presents significant challenges. The behavioural marker system must capture the context in which the assessment is made (e.g., crew dynamics and experience, operating environment, operational complexity). For example, in a team endeavour, the behaviour of one crew member can be adversely or positively impacted by another, resulting in a substandard or inflated performance rating. Behavioural marker systems should be designed to detect and record such effects.

TRAINING

10. What are prerequisites to be a trainer for a behavioural marker course?

Qualifications required of the persons who will deliver a formal course to train, calibrate, and qualify raters (evaluators) using the behavioural marker system:

- Commitment to human factor's principles
- Domain knowledge
- Formal training in applicable aspects of human factors or non-technical skills (e.g., Crew Resource Management)
- Formal training in the use and limitations of performance rating systems
- Formal training in the use of the specific behavioural marker system

11. What are prerequisites for evaluators using a behavioural marker system?

Entry requirements for personnel who will serve as evaluators:

- Commitment to human factor's principles
- Domain knowledge
- Formal training in applicable aspects of Human Factors or non-technical skills (e.g., Crew Resource Management)

12. What are necessary qualifications of evaluators?

- Complete initial training in behavioural marker systems
- Formal assessment as competent and calibrated following behavioural marker system-training in classroom
- Calibration in operational environment (e.g., training, simulator, work environment)
- Periodic re-calibration for continuing use of the behavioural marker system

FREQUENTLY ASKED QUESTIONS ABOUT BEHAVIOURAL MARKER SYSTEMS

TRAINING

13. What should the content of behavioural marker system training be?

- Make explicit goals for use of the behavioural marker system
- Explain the design of the behavioural marker system, as well as content and guidelines for its use
- Review main sources of rater biases (e.g., hindsight, halo, recency, primacy) with techniques to be used for minimisation
- Present the concept of inter-rater reliability and the methods to be used to maximise it
- Illustrate and define each point of the rating scale and different levels of situational complexity with video examples, discussions, and hands-on exercises
- Provide practical training with multiple examples
- Include calibration with iterative feedback on inter-rater reliability score
- Teach debriefing skills as appropriate
- Conclude with a formal assessment of rater competence

14. How should a behavioural marker system training be structured?

- Minimum two consecutive days training
- An ideal group size of 8-12 people
- Training follow-up (e.g., meetings, feedback via telephone) after use of behavioural marker system in operational setting
- Training ideally utilises video examples from the organisation.

15. What training and calibration materials should be used?

- Videotapes of scenarios with professional sound and visual quality
  - Demonstrating various levels of performance
  - Showing all behavioural markers in scenarios illustrating various environmental conditions and complexity
  - Depicting increasing lengths of segments with training progress (e.g., from 2 minute vignettes of a specific behavioural marker to an entire flight/surgical procedure)
- Information about the background of the behavioural marker system with full reference documentation

FREQUENTLY ASKED QUESTIONS ABOUT BEHAVIOURAL MARKER SYSTEMS

REGULATORY AND RESEARCH ISSUES

16. What are regulatory issues regarding the use of behavioural marker systems?

- The rationale for employing a behavioural marker system in any domain is to improve levels of safety and to facilitate attainment of the highest possible levels of performance.
- A partnership between the operators and regulatory authorities is needed to achieve equitable assessment of non-technical skills, especially when a pass/ fail criterion is mandated.
- Regulators should move cautiously when initiating formal assessment of non-technical skills.

17. What are research issues regarding the use of behavioural marker systems?

By nature, behavioural marker systems are not static, but must be continually evolved or refined in response to changing operational circumstances (e.g., development of equipment) and increased understanding of human factors issues in the domain. The following list, which is not exhaustive, specifies research topics where empirical evidence is either lacking or incomplete and systematic research should prove highly beneficial:

- Developing empirical evidence for the relative merits of global vs. phase or event-specific ratings and individual vs. team ratings
- Defining context effects on crew behaviour and developing a systematic system of integrating these measures with behavioural markers to provide a more comprehensive system
- Investigating the distribution of ratings of markers taken in different data collection environments (i.e., training including technical and non-jeopardy, full mission simulation (LOFT), non-jeopardy assessment of system performance (LOSA), formal evaluations in both line operations, and recurrent proficiency checks)
- Integrating knowledge from incident analyses, especially coping/recovery strategies and translating them into behavioural markers
- Providing practical guidance for the transfer of behavioural marker systems and/or their components across domains and cultures (national, professional, and organisational)

FREQUENTLY ASKED QUESTIONS ABOUT BEHAVIOURAL MARKER SYSTEMS

CONCLUSION

Conclusion

- Behavioural marker systems have demonstrated value for training, understanding of performance in high risk environments, and research into safety and human factors.
- Behavioural marker systems can contribute to safety and quality in other work environments, as well as in high risk settings.
- Concepts are continuously evolving as a result of co-operation between practitioners and researchers.
- Researchers, practitioners, and regulatory authorities must work congruently in order to realise the ultimate goal of improved safety.

BIBLIOGRAPHY

Andlauer, E. & the JARTEL group (in prep.). Joint Aviation Requirements - Translation and Elaboration. JARTEL Project Report to DG-TREN European Commission. Paris: Sofreavia.

Antersijn, P. & Verhoef, M. (1994). Assessment of non-technical skills: is it possible? In N. McDonald, N. Johnston & R. Fuller (Eds.), Application of Psychology to the Aviation System. Avebury: Aldershot.

Avermaete, J. & Kruijsen, E. (Eds.) (1998). NOTECHS. The evaluation of non-technical skills of multi-pilot aircrew in relation to the JAR-FCL requirements. Final Report NLR-CR-98443. Amsterdam National Aerospace Laboratory (NLR).

Fletcher, G., McGeorge, P, Flin, R., Glavin, R. & Maran, N. (in press). The role of non-technical skills in anaesthesia. British Journal of Anaesthesia.

Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N. + Patey, R. (2001). Development of a Behavioural Marker System for Anaesthetists' Non-Technical Skills (ANTS). University of Aberdeen Grant Report RDNES/991/C for SPMDE.

Flin, R., Goeters, K.-M., Hoermann, J. & Martin, L. (1998). A generic structure of non-technical skills for training and assessment. Paper presented at 23rd Conference of the European Association for Aviation psychology, Vienna, 14-18 September.

Flin, R. & Martin, L. (2001). Behavioural markers for Crew Resource Management: A survey of current practice. International Journal of Aviation Psychology, 11, 95-118.

Hart, S.G & Staveland, L.E. (1988). Development of NASA-TLX (Task Load indeX): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), Human Mental Workload. Amsterdam: Elsevier.

Helmreich, R.L., Butler, R.E., Taggart, W.R., & Wilhelm, J.A. (1995). Behavioral markers in accidents and incidents: Reference list. NASA/UT/FAA Technical Report 95-1. Austin, TX: The University of Texas.

Helmreich, R.L., Chidester, T.R., Foushee, H.C., Gregorich, S.E., & Wilhelm, J.A. (1990, May). How effective is Cockpit Resource Management training? Issues in evaluating the impact of programs to enhance crew coordination. Flight Safety Digest, 9(5), 1-17. Arlington, VA: Flight Safety Foundation.

Helmreich, R.L., & Foushee, H.C. (1993). Why Crew Resource Management? Empirical and theoretical bases of human factors training in aviation. In E. Wiener, B. Kanki, & R. Helmreich (Eds.), Cockpit Resource Management (pp. 3-45). San Diego, CA: Academic Press.

BIBLIOGRAPHY

Helmreich, R.L., Foushee, H.C., Benson, R., & Russini, R. (1986). Cockpit management attitudes: Exploring the attitude-performance linkage. Aviation, Space and Environmental Medicine, 57, 1198-1200.

Helmreich, R.L., Kello, J.E., Chidester, T.R., Wilhelm, J.A., & Gregorich, S.E. (1990). Maximizing the operational impact of Line Oriented Flight Training (LOFT): Lessons from initial observations. NASA/University of Texas Technical Report 90-1.

Helmreich, R.L., Klinect, J.R., & Wilhelm, J.A. (in press). System safety and threat and error management: The line operations safety audit (LOSA). In Proceedings of the Eleventh International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Helmreich, R.L., Wilhelm, J.A., Gregorich, S.E., & Chidester, T.R. (1990). Preliminary results from the evaluation of Cockpit Resource Management training: Performance ratings of flightcrews. Aviation, Space, and Environmental Medicine, 61(6), 576-579.

Hoermann, H.-J. (2001) Cultural variation of perceptions of crew behaviour in multi-pilot aircraft. Le Travail Humain, 64 (3), 247-268.

Klampfer, B., Häusler, R., Fahnenbruck, G. & Naef, W. (in press). Group Interaction in High Risk Environment – Outline of a Study. In M.Cook (Ed.), Proceedings of the 24th Conference of the European Association of Aviation Psychology, Crieff, Scotland , September 4 - 8, 2000. Aldershot, UK: Ashgate.

O'Connor, P., Hoermann, J., Lodge, M., Flin, R. et al. (in press). Training non-technical skills: A European perspective. International Journal of Aviation Psychology.

O'Connor, P., Flin, R. & O'Dea, A. (2001). Team skills for nuclear power plant operating staff. Technical Report to Nuclear Industry Management Committee. Industrial Psychology Group, University of Aberdeen.

Salas, E., Bowers, C. & Edens, E. (Eds.) (2001) Improving teamwork in organizations. Applications of resource management training. Mahwah, N.J.: LEA.

Sexton, J.B., Klinect, J.R., & Helmreich, R.L. (in press) The link between safety attitudes and observed performance in flight operations. Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

APPENDIX 1: THE GIHRE-AVIATION PROJECT

Background

Group Interaction in High Risk Environments (GIHRE) is the name of an interdisciplinary project that was launched by the Gottlieb Daimler and Karl Benz Foundation in 1998. Of main interest is the management of high workload situations by small groups of professionals working in high risk environments, such as cockpit crews, surgery teams, and nuclear power plant control teams.

As Werner Naef, Head of Human Aspects Development, Swissair, was asked to head the subproject GIHRE-aviation, it was the opportunity for Swissair to investigate questions that could give an answer to practical concerns within the area of CRM assessment. Two psychologists with aviation background and one pilot completed the project team. A simulator study was then designed to test behavioural markers for their use in situations with high workload. The two behavioural marker systems selected for this purpose were the NOTECHS (non-technical skills) Markers and University of Texas (UT) Behavioural Markers. NOTECHS was already being used for some training purposes within Swissair. With the support of Robert Helmreich, of the head of another GIHRE subproject, the use of the UT markers was agreed upon.

Aims & Questions

The main objective of the GIHRE-aviation project is the validation of existing behavioural markers for CRM assessment under conditions of high workload. In essence, this involves the comparison of the American and the European marker sets with regard to several criteria in the simulator environment. Additional questions to be answered through the project's research are: Which behavioural markers differentiate best between outstanding and substandard crews under high workload? Are good team players always good team players or is team performance more situation dependant? Is there a relationship between CRM behaviour, technical performance and errors, and if so, what is it?

Data

The database for analysis consists of videotaped simulator sessions from the recurrent training of Swissair A320 and Lufthansa CityLine Canadair Jet fleets. During the simulator sessions, three predefined scenarios with different workload levels are recorded. The variation between the three scenarios regarding mental and manual workload is verified with a subjective workload measurement – the NASA Task Load Index (TLX: Hart & Staveland, 1988). The TLX is filled out right after the simulator session, before the debriefing takes place. Additionally, pilots answer short questionnaires that include biographical data and a self-assessment. There is also an expert judgement from the simulator instructors, who rate different aspects of team performance. The video is recorded so that primary flight parameters such as speed, altitude, heading, and power setting can be seen at the bottom of the screen.

The video data collected during the simulator sessions are then analysed by the GIHRE-aviation team with respect to the NOTECHS and UT behavioural markers. Flight parameters and the expert ratings from flight instructors serve as reference measures for technical performance.

Status of the Project

To date, 80 Swissair crews and 25 crews from Lufthansa CityLine have been video recorded. A sub-sample of 23 videos from Swissair crews have been analysed; video analyses for the remaining tapes are still in progress and will be finished by the end of 2001.

Preliminary results from the 23 analysed Swissair crews show that the CRM performance of the same crew varies between situations. It is too early to comment on the ability of single markers to differentiate between good and poor performing crews. What can be said is that there are some markers within each of the two behavioural marker systems that can be observed more often than others. Some markers also tend to show less variation in their ratings than other markers. Regarding the two behavioural marker systems (NOTECHS and UT Markers), it can be concluded that they complement each other, measuring a variety of CRM related behaviours from different perspectives not covered by one system alone.

It is anticipated that following analyses of the full sample, clear statements will be possible regarding the strengths and limitations of single behavioural markers as well as behavioural marker systems in the flight simulator environment.

APPENDIX 2:
A BRIEF HISTORY OF THE ORIGINS AND EVOLUTION OF UT BEHAVIOURAL MARKERS

The original behavioural marker system in the U.S. originated in the University of Texas Human Factors Research Project (then called the NASA/UT Project) in the late 1980s. There were two goals associated with the effort: the first was to evaluate the effectiveness of CRM training as measured by observable behaviours, while the second was to aid in defining the scope of CRM programmes. The first manual to assist check airmen and evaluators in assessing the interpersonal component of flying was issued by NASA/UT in 1987. Originally, ratings of crew performance were made by observers assessing a complete flight from initial briefing to landing, taxi-in, and shutdown of engines.

The first set of behavioural markers was included by the Federal Aviation Administration as an appendix to its Advisory Circular on CRM (AC-150A). Development of the markers was supported by a grant from the FAA. Systematic use of the markers grew as airlines enhanced assessment of crew performance and as the University of Texas project began collecting systematic data on all aspects of an airline's operations in a programme known as the Line Operations Safety Audit (LOSA). The markers themselves were incorporated in a form for systematic observations known as the Line/LOS Checklist (LOS refers to line operational or full mission simulation). As experience and the database of observations grew, it became apparent that there was significant variability in crew behaviour during flights that needed to be captured. Accordingly, the form was modified to assess the markers for each phase of flight.

In 1995, a validation of the markers was undertaken by classifying their impact (positive and negative) in analyses of aviation accidents and incidents. The results of the analysis provided strong support for the utility of the markers as indicators of crew performance and their value as components of CRM training.

LOSA evolved over time from a sole assessment of the behavioural markers to a focus on threat and error management. In this iteration (now reflected in the 9th generation of the data collection instrument), threats and errors are classified and their management assessed along with a greatly reduced set of behavioural markers. The new focus on threat and error management provided hard, empirical criteria against which to pit the markers. In this process, a number of overlapping markers were dropped to yield a smaller, but highly influential list. These are shown, along with the phase of flight in which collected, in the following section.

Training for LOSA, including the behavioural markers as well as classification of threats and errors, takes two full days and is similar to that recommended for using the markers alone.

APPENDIX 2:
UNIVERSITY OF TEXAS BEHAVIOURAL MARKERS RATING SCALE

The markers listed below are used in Line Operations Safety Audits, non-jeopardy observations of crews conducting normal line flights. Each of these markers has been validated as relating to either threat and error avoidance or management. With the exception of two global ratings, specific markers are rated (if observed) during particular phases of flight. Following is a list of currently used markers showing phase where rated, followed by the ratings for each phase of flight:

Key to Phase: P = Pre-departure/Taxi; T = Takeoff /Climb; D = Descent/Approach/Land; G = Global

			Phase
SOP BRIEFING	The required briefing was interactive and operationally thorough	- Concise, not rushed, and met SOP requirements - Bottom lines were established	P-D
PLANS STATED	Operational plans and decisions were communicated and acknowledged	- Shared understanding about plans - "Everybody on the same page"	P-D
WORKLOAD ASSIGNMENT	Roles and responsibilities were defined for normal and non-normal situations	- Workload assignments were communicated and acknowledged	P-D
CONTINGENCY MANAGEMENT	Crew members developed effective strategies to manage threats to safety	- Threats and their consequences were anticipated - Used all available resources to manage threats	P-D
MONITOR / CROSSCHECK	Crew members actively monitored and cross-checked systems and other crew members	- Aircraft position, settings, and crew actions were verified	P-T-D
WORKLOAD MANAGEMENT	Operational tasks were prioritized and properly managed to handle primary flight duties	- Avoided task fixation - Did not allow work overload	P-T-D
VIGILANCE	Crew members remained alert of the environment and position of the aircraft	- Crew members maintained situational awareness	P-T-D
AUTOMATION MANAGEMENT	Automation was properly managed to balance situational and/or workload requirements	- Automation setup was briefed to other members - Effective recovery techniques from automation anomalies	P-T-D
EVALUATION OF PLANS	Existing plans were reviewed and modified when necessary	- Crew decisions and actions were openly analyzed to make sure the existing plan was the best plan	P-T
INQUIRY	Crew members asked questions to investigate and/or clarify current plans of action	- Crew members not afraid to express a lack of knowledge - "Nothing taken for granted" attitude	P-T
ASSERTIVENESS	Crew members stated critical information and/or solutions with appropriate persistence	- Crew members spoke up without hesitation	P-T
COMMUNICATION ENVIRONMENT	Environment for open communication was established and maintained	- Good cross talk – flow of information was fluid, clear, and direct	G
LEADERSHIP	Captain showed leadership and coordinated flight deck activities	- In command, decisive, and encouraged crew participation	G

1 = Poor	2 = Marginal	3 = Good	4 = Outstanding
Observed performance had safety implications	Observed performance was barely adequate	Observed performance was effective	Observed performance was truly noteworthy

APPENDIX 3:
THE DEVELOPMENT OF THE NOTECHS BEHAVIOURAL MARKERS

The European Joint Aviation Requirements (JAR) require the training and assessment of pilots' CRM skills. JAR Ops NPA 16 states: "The flight crew must be assessed on their CRM skills in accordance with a methodology acceptable to the Authority and published in the Operations Manual. The purpose of such an assessment is to: Provide feedback to the individual and serve to identify retraining; and be used to improve the CRM training system". CRM skills can also be called non-technical skills. These refer to a flight crew member's behaviours in the cockpit not directly related to aircraft control, system management, and standard operating procedures.

In 1996, the JAA Research Committee on Human Factors initiated a project that was sponsored by four European Civil Aviation Authorities (Germany, France, Netherlands, UK). A research consortium, consisting of members from DLR (Germany), IMASSA (France), NLR (Netherlands), and University of Aberdeen (UK), was established to work on what was called the NOTECHS (Non-Technical Skills) project. This group was required to identify or develop a feasible and efficient methodology for assessing pilots` non-technical skills. The design requirements were (i) that the system was to be used to assess the skills of an individual pilot, rather than a crew, and (ii) it was to be suitable for use across Europe, by both large and small operators. After reviewing existing methods it became apparent, for various reasons (e.g., crew-based, fleet specific, or too complex), that none of these systems met the design requirements and therefore they could not be taken as an Acceptable Means of Compliance (AMC) under the scope of the JAR. Moreover, none of them provided a suitable basis for simple amendment, although particular attention was paid to two of the principal frameworks, namely the KLM WILSC/ SHAPE systems (see Antersijn & Verhoef, 1994; Avermaete & Kruijsen, 1998) and the NASA UT Line/LOS Checklist system (LLC version 4) (Helmreich et al., 1995). Therefore, the research group, with the assistance of training captains from KLM, designed a prototype behavioural marker system for rating non-technical skills, which was called NOTECHS.

The development of the NOTECHS system consisted of: (i) the review of existing systems to evaluate proficiency in non-technical skills (see also Flin & Martin, 1998); (ii) a literature search for relevant research findings relating to key categories of non-technical skills; (iii) extended discussions with subject matter experts at NOTECHS working group meetings. The following design criteria were used to guide the final choice of components and descriptor terms: a) the system should contain the minimum number of categories and elements in order to encompass the critical behaviours; b) the basic categories and elements should be formulated with minimum overlap; c) the terminology

The JARTEL project

In 1998, a European project team was established to work on the JARTEL project . This team was funded by the European Commission (DG TREN) and consisted of the following partners: Alitalia, British Airways, Airbus, DERA (UK), DLR (G), IMASSA (F), NLR (N), Sofreavia (F) and University of Aberdeen (UK), several of whom had been involved in the NOTECHS project. The basic aim of the JARTEL project was to conduct initial tests of the NOTECHS behavioural marker system to ascertain whether it was a) reliable, b) usable, and c) culturally robust across European operators. This project began with a literature review to determine the main cultural clusters relating to flight deck crews' behaviour patterns in Europe. This was followed by an experimental study with 105 training captains from larger and smaller operators located in the five main cultural clusters using the NOTECHS system (see Hoermann, 2001; O'Connor et al., in press). An operational study involving a number of airlines has just been completed. The final project report will be available by the end of 2001 (Andlauer et al., in prep.).

should reflect everyday language for behaviour, rather than psychological jargon; d) the skills listed at the behavioural level should be directly observable in the case of social skills or inferable from crew interaction in the case of cognitive skills. Details of the development process and discussion of the legal and methodological background of the assessment of flight crew members' non-technical skills can be found in Avermaete and Kruijsen (1998).

The resulting structure of NOTECHS comprises four categories: Cooperation, Leadership & Managerial Skills, Situation Awareness, Decision Making. These four primary categories effectively subdivide into two social skills categories (Cooperation; Leadership & Managerial) and two cognitive skills categories (Situation Awareness; Decision Making). In relation to the four categories, 15 elements were identified. For each element, a number of positive and negative exemplar behaviours were included. These were phrased as generic behaviours (e.g., closes loop for communications) rather than specific behaviours (e.g., reads back to ATC) to give an indication of type, and to avoid designating particular actions that should be observed.

Finally, a set of NOTECHS Principles was established, which should be adhered to when the system is used.

- Evaluations based on observable behaviours
- Need for technical consequences
- Evaluations based on repeatedly shown behaviour patterns
- Scale should allow for ratings of acceptable to unacceptable behaviours
- Explanation required if unacceptable category rating is given.

The NOTECHS system requires a minimum of two full days of specialist training and this should meet the recommendations in the guidelines above. Further information on the development and evaluation of the NOTECHS system can be found in the Joint Aviation Requirements – Translation and Elaboration of Legislation (JARTEL) project report (see above).

APPENDIX 3:
THE NOTECHS BEHAVIOURAL MARKERS
THE NOTECHS RATING SCALE

Categories	Elements	Example Behaviours
COOPERATION	Team building and maintaining	- Establishes atmosphere for open communication and participation
	Considering others	- Takes condition of other crew members into account
	Supporting others	- Helps other crew members in demanding situation
	Conflict solving	- Concentrates on what is right rather than who is right
LEADERSHIP & MANAGERIAL SKILLS	Use of authority and assertiveness	- Takes initiative to ensure involvement and task completion
	Maintaining standards	- Intervenes if task completion deviates from standards
	Planning and co-ordinating	- Clearly states intentions and goals
	Workload management	- Allocates enough time to complete tasks
SITUATION AWARENESS	System awareness	- Monitors and reports changes in system's states
	Environmental awareness	- Collects information about the environment
	Anticipation	- Identifies possible future problems
DECISION MAKING	Problem definition / diagnosis	- Reviews causal factors with other crew members
	Option generation	- States alternative courses of action - Asks other crew member for options
	Risk assessment / Option choice	- Considers and shares risks of alternative courses of action
	Outcome review	- Checks outcome against plan

Very Poor	Poor	Acceptable	Good	Very Good
Observed behaviour directly endangers flight safety	Observed behaviour in other conditions could endanger flight safety	Observed behaviour does not endanger flight safety but needs improvement	Observed behaviour enhances flight safety	Observed behaviour optimally enhances flight safety and could serve as an example for other pilots

APPENDIX 4:
ILLUSTRATIVE COMPARISON OF NOTECHS ELEMENTS AND UT MARKERS

Table 1:
Comparison of NOTECHS categories and elements with the UT markers

While the University of Texas (UT) and the NOTECHS behavioural marker systems were designed for different purposes, the fundamental behavioural components of the two systems are similar. The tables below illustrate that many of the elements in the NOTECHS system are comparable to markers in the UT system. As explained above, the designers of the NOTECHS system took the core components of an earlier version (LLC4) of the UT markers into account when selecting the principal elements in NOTECHS.

NOTECHS		University of Texas
Categories	Elements	Markers
COOPERATION	Team building and maintaining	Communication Environment
	Considering others	–
	Supporting others	–
	Conflict solving	–
LEADERSHIP & MANAGERIAL SKILLS	Use of authority and assertiveness	Assertiveness Leadership Inquiry
	Maintaining standards	SOP briefing
	Planning and co-ordinating	Plans stated SOP briefing Evaluation of plans Leadership Workload assignment
	Workload management	Workload assignment Workload management Automation management
SITUATION AWARENESS	System awareness	Monitor / Crosscheck Automation management
	Environmental awareness	Vigilance
	Anticipation (awareness of time)	Contingency management
DECISION MAKING	Problem definition / diagnosis	Inquiry
	Option generation	Inquiry
	Risk assessment / option selection	Contingency management
	Outcome review	Evaluation of plans

APPENDIX 4:
ILLUSTRATIVE COMPARISON OF NOTECHS ELEMENTS AND UT MARKERS

Table 2:
Comparison of UT categories and markers with the NOTECHS element

University of Texas		NOTECHS
Categories	Markers	Elements
PLANNING	SOP briefing	Maintaining standards Planning and co-ordinating
	Plans stated	Planning and co-ordinating
	Workload assignment	Workload management Planning and co-ordinating
	Contingency management	Anticipation (awareness of time) Risk assessment / option selection
EXECUTION	Monitor / Crosscheck	System awareness
	Workload management	Workload management
	Vigilance	Environmental awareness
	Automation management	Workload management System awareness
REVIEW / MODIFY	Evaluation of plans	Planning and co-ordinating Outcome review
	Inquiry	Problem definition Option generation
	Assertiveness	Use of authority and assertiveness
	Communication environment	Team building and maintaining
	Leadership	Use of authority and assertiveness Planning and co-ordinating

APPENDIX 5:
BIOGRAPHIES OF PARTICIPANTS

Andrea Amacher is a student at the Institute of Work and Organisational Psychology at the University of Bern, Switzerland. She became familiar with the field of aviation psychology through her work in the Swissair Pilot Selection Division. Since April 2001, she has been part of the GIHRE-aviation project. In her master thesis she is investigating the relationship between cockpit-crew performance and error in high workload situations.

Peter Dieckmann is a research assistant with the Department of Psychology of the University of Heidelberg. Currently he is involved in a project dealing with the optimisation and evaluation of simulator training courses in Anaesthesia Crisis Resource Management (ACRM) with the university's own anaesthesia simulator. His research interests include the use of simulators for training and research concerning operator and team reliability, especially conditions for effective simulator training. He did a survey study comparing simulator training in six different domains (e.g., aviation and anaesthesia).

Paul Field is an Airline Pilot with British Airways. He currently manages the design of Human Factors and Non-technical Training for the airline's flightcrew, and co-ordinates the airline's contribution to the Enhanced Safety through Situation Awareness Integration in training (ESSAI) research project. ESSAI is sponsored by the European Union, and seeks to develop training strategies that specifically enhance Situation Awareness and 'threat and error management' on the flightdeck. He graduated from Newcastle University in 1978 with a BSc in Physics, and then flew fast-jets with the Royal Air Force before joining the British Airways in 1987. He is also a member of the Royal Aeronautical Society's CRM Advisory Panel.

Georgina Fletcher is a research fellow in the Industrial Psychology Group at the University of Aberdeen in Scotland. She has a BSc in Psychology from Bristol University and an MSc in Ergonomics from University College London. From 1994 to 1999 she worked at the Defence Evaluation and Research Agency on a variety of human factors research projects, mainly in the civil aviation field, including: the evaluation of pilots' non-technical skills and investigating training requirements for pilots converting to highly automated 'glass cockpit' aircraft. Since then she has been working on a project funded by the Scottish Council for Postgraduate Medical and Dental Education (SCPMDE) to develop and evaluate a behavioural marker system for Anaesthetists' Non-Technical Skills. Once validated, this system will be used to support simulator-based training and assessment for anaesthetists in Scotland. This work is also the subject of her PhD thesis.

Rhona Flin is Professor of Applied Psychology in the Department of Psychology at the University of Aberdeen. She holds BSc and PhD degrees in Psychology from the University of Aberdeen, is a

APPENDIX 5: BIOGRAPHIES OF PARTICIPANTS

Chartered Psychologist, a Fellow of the British Psychological Society, and a Fellow of the Royal Society of Edinburgh. With her colleagues in the Industrial Psychology Group, she studies safety and emergency management in high reliability industries. Current projects include the development of a behavioural marker system for airline pilots' non-technical skills (EC); CRM evaluation (CAA); human factors and safety in offshore oil operations (HSE/ oil industry); decision making in emergency management (nuclear industry); team skills (nuclear industry); management impact on safety climate (oil/ power industries), and the development of a behavioural marker system for anaesthetists' non-technical skills (SCPMDE).

Ruth Haeusler is a research and teaching assistant at the Department of Work and Organisational Psychology of the University of Bern, Switzerland. She received her master's degree in Psychology at the University of Bern in 2000. For four years she has been working in the field of aviation psychology, starting with her master thesis on the measurement of CRM behaviour at Swissair. She is currently working in the GIHRE-aviation project. Another project links her to the Department of Anaesthesiology at the University Hospital of Basel, in which a CRM measurement tool for full surgical teams is being developed and validated. Subject of her doctoral thesis are work strategies of pilots and their effect on performance in high workload situations.

Robert L. Helmreich, PhD, FRAeS is Professor of Psychology at The University of Texas at Austin. He is Director of the University of Texas Aerospace Crew Research Project. He obtained his BS degree in Psychology, Sociology, Anthropology, and Biology at Yale University and MS and PhD in personality and social psychology. The project investigates issues in crew selection, training, and performance in both aviation and space environments. The project is also investigating human factors issues in medicine using survey and observational methodologies with a focus on the operating theatre, emergency room, and Intensive Care Unit. He is a Fellow of the Royal Aeronautical Society, American Psychological Association, and American Psychological Society. He received the Flight Safety Foundation Distinguished Contribution to Aviation Safety Award for 1994 and the 1994 Laurels from Aviation Week and Space Technology for his role in the development of CRM. In 1997, he received the David S. Sheridan Award for contributions to patient safety in medicine.

Barbara Klampfer is a research assistant at the Swiss Federal Institute of Technology at Zurich. She received her master's degree in Psychology at the University of Salzburg in 1995, where she also did postgraduate work on Training Conception and Methodology. Since 1997, she holds the Private Pilot License. From 1997 to 1999, she worked at a project dealing with incident analysis based on in-flight data under grants from Swissair Flight Safety Department and the Swiss Reinsurance Company. Out of this, her doctoral thesis focuses on the influence of automation on minor incidents. Since 1999, she has been working in the GIHRE-aviation project.

APPENDIX 5:
BIOGRAPHIES OF PARTICIPANTS

Kristina Lauche, PhD, is a lecturer in Industrial and Organisational Psychology at the University of Aberdeen, Scotland. She received her master's degree in Psychology from the Free University Berlin, Germany, and her PhD from the University of Potsdam, Germany. She has worked as a researcher in applied industrial psychology at the University of Munich on quality management and training. From 1997 to 2001, she worked as a research fellow at the Swiss Federal Institute of Technology (ETH) Zurich on computer supported co-operative work and multidisciplinary teams. Her main areas of research are innovation and heedful interrelating in teams.

J. Bryan Sexton is a doctoral candidate at The University of Texas at Austin. He received both his bachelor's and master's degrees in Psychology at The University of Texas. In 1995, he spent a year in Switzerland as a visiting scholar at the University Hospital of Basel. His work there examined human factors in the operating room through surveys, observational studies, and the development of a high-fidelity operating room simulator for training full surgical teams. Since 1994, he has worked at The University of Texas Human Factors Research Project under grants from NASA, the Federal Aviation Administration, the Agency for Healthcare Research and Quality, the Swiss National Science Foundation, and the Gottlieb Daimler and Karl Benz Foundation. This work has taken him from the studying pilot selection and flight safety in the cockpit to investigations of patient safety and safety culture in the intensive care unit. He currently serves as Human Factors Advisor to The Society for Human Performance in Extreme Environments. His dissertation links human factors to patient outcomes in a national sample of 106 ICUs in the United Kingdom.

Sven Staender, MD is a professional anaesthetist and intensive care physician. He graduated from medical school in 1987 and started his medical specialisation training at the University of Basel in 1987, where he became a faculty member in 1993. He left the University Hospital in Basel in 1999 to become head of the department of anaesthesia and intensive care at the regional hospital Maennedorf/Zurich in Switzerland. Since the start of human factors activities at the University of Basel in 1991 (full scale simulator training etc.) he has been part of the team developing a critical incident reporting system for anaesthesiologists. Today he is responsible for the national incident reporting system in anaesthesiology in Switzerland.

Impressum

Group Interaction in High Risk Environments (GIHRE)
Berlin, November 2001

Project Head:
Prof. Dr. Rainer Dietrich

Co-ordination:
Kateri Jochum

Humboldt Universität zu Berlin
(Sitz: Schützenstr. 21)
Unter den Linden 6
10099 Berlin
Tel. +49-30-20 939-673 and 20 939-772
Fax: +49-30-20 939-729
URL: http://www2.hu-berlin.de/GIHRE

Editors:
Barbara Klampfer
Kateri Jochum

Artwork:
Karsten Reuß